UNIVERSITY OF CALIFORNIA

Santa Barbara

Assessing the Factorial Validity, Measurement Invariance, and Latent Mean Differences of a Second-Order, Multidimensional Model of Academic and Social College Course Engagement: A Comparison Across Course Format, Ethnic Groups, and Economic Status

A dissertation submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

EDUCATION

by

Juan Emilio Espinosa

Committee in Charge:

Professor John T. Yun, Michigan State University (Co-Chair)

Professor Rebeca Mireles-Rios (Co-Chair)

Professor Cynthia Hudley

Professor Richard P. Duràn

December 2016

The Dissertation of
Juan Emilio Espinosa is approved:

Richard Duràn

Cynthia Hudley

Rebeca Mireles-Rios, Committee Co-Chair

John T. Yun, Committee Co-Chair

December 2016

Assessing the Factorial Validity, Measurement Invariance, and Latent Mean Differences of a Second-Order, Multidimensional Model of Academic and Social College Course Engagement: A Comparison Across Course Formats, Ethnic Groups, and Economic Status

DEDICATION

For my mother, Irma Sevilla Rosas, and my grandmother, Rebecca Serna Espinosa.

Even after your departure, you bless me with your presence, kind, and grace;
shine a glimmer of hope, inspiration, and reassurance; and
impart wisdom, guidance, and love.

Until we meet again –

ACKNOWLEDGEMENTS

I would like to begin by expressing my sincere gratitude and appreciation to the members of my all-star committee, Drs. Richard Duràn, Cynthia Hudley, Rebecca Mireles-Rios, and John Yun. Richard, your intellect shines through each of our conversations and meetings. I strive to amass your level of knowledge. Cynthia, you have taught me so much about educational psychology and achievement motivation. Your influence has spanned into my personal life and the lives of my family. Rebecca, it was truly a blessing that you agreed to join my committee. Your dedication to students shines bright, and I am confident it will not pale with time. John, I still recall the first time you called me, and I am so glad that you chose not to immediately rescind my acceptance. All jokes aside, thank you for exposing me to the field of evaluation, serving as my advisor and mentor, bringing me aboard the UCEC team, and dedicating your own personal time to ensure that I completed this program. My graduate training began during our weekly JPME meetings; I can't thank you enough for all of your support. Of course, there is no way I could leave out the better half of my favorite couple in academia—Dr. Patricia Marin. (One day I'm bound to meet another couple, :). Patricia, I recall mentioning my interest in working with you during my statement of purpose. Although you were unable to serve on my committee, my wish came true. More importantly, I gained a life-long mentor and friend. I have learned so much from your style of leadership. My family and I thank you and John for all that you two have done for me.

At San Jose State University, I would like thank Valerie Chapman for forcing me to enroll in the Success as Transfers program. That course taught me how to be a successful student. Thank you, Michael Randle, for exposing me to student services

and program coordination. My experience working for the Success as Transfers program played a major role in my decision to dedicate my life to educational service. Sandi Douglas, Edwin Hunt, Drs. John Jabagchourian, Meekyung Han, and Emily Bruce, thank you for your support. There are so many people that have impacted my life and provided unwavering support throughout my educational journey and my quest to support others on this same path. All of you are appreciated.

I am forever indebted to my family for encouraging me to stay the path and delay gratification. I must thank my Uncle Gil Espinosa, Tia Estela Rosas, and Tia Graciela Rosas. You have supported my educational pursuits throughout my entire life. Tia Virginia Rosas, thank you for the sleepless nights that you endured helping me apply to graduate school. Gabrielle E. Torres, thank you for being the great, caring person you are. I can't forget my Tia Armida Rosas and Uncle David Cotti. I've learned so much from you. I can't thank you two enough for all that you have done for me. I have so many wonderful family members, on both sides of my family, that have provided me with unconditional love. I love each and every one of you, and I wanted to include all of you, but that would have been another dissertation in itself.

Lastly, I would like to thank my immediate family members. I am thankful for my father, Juan Antonio Espinosa. You taught me at young age that anything is attainable with hard work and dedication. That's been the most valuable lesson that I've learned. I can't thank my brother enough, Carlos Castorena, I've admired you since I was a kid, and I continue to look up to you to this day. Thank you, Emilio Tomas Castorena and Carlos Castorena, Jr. You boys are up next! Mom, I made sure to complete my program because of you. I miss and love you so much.

VITA OF JUAN EMILIO ESPINOSA

December 2016

EDUCATIONAL HISTORY

UNIVERSITY OF CALIFORNIA, SANTA BARBARA

Doctor of Philosophy in Education 2016
Master of Arts in Education 2013
Emphasis: Educational Leadership & Organizations

SAN JOSÉ STATE UNIVERSITY

Bachelor of Arts in Social Work 2010
Minor in Child and Adolescent Development

AREAS OF EXPERTISE

Program Planning & Development	Student Services & Success
Educational Assessment & Evaluation	Online Education
Educational Psychology	Covariance Analysis

RESEARCH EXPERIENCE

UNIVERSITY OF CALIFORNIA, DAVIS

Department of Public Health Sciences 2016
Evaluation Associate

UNIVERSITY OF CALIFORNIA, SANTA BARBARA

Office of Education Partnerships 2013 – 2014
Graduate Student Evaluator

The Institute for Social, Behavioral and Economic Research 2013 – 2014
Graduate Student Assistant

University of California Educational Evaluation Center 2010 – 2013
Graduate Student Researcher

STUDENT AFFAIRS EXPERIENCE

UNIVERSITY OF CALIFORNIA, SANTA BARBARA

The Institute for Social, Behavioral and Economic Research 2013 – 2014
Graduate Student Assistant

UC-HBCU Initiative: UCSB-FAMU Partnership 2011 – 2013
Graduate Student Mentor

Office of Academic Preparation 2010 – 2011
Graduate Student Assistant

SAN JOSÉ STATE UNIVERSITY

San José State University Research Foundation
Peer Advisor Coordinator 2009 – 2010
Peer Advisor 2007 – 2009

TECHNICAL REPORTS

Espinosa, J. E., Cassady, D. L. (2016). Prevention First (1305) enhanced evaluation report, Year 3: Increase use of team-based care in health systems. University of California, Davis: Public Health Sciences Department.

Treiber, J., & **Espinosa, J. E.,** Cassady, D. L. (2016). Prevention First (1305) enhanced evaluation report, Year 3: Implement quality physical education and physical activity. University of California, Davis: Public Health Sciences Department.

Yun, J. T., Grimes, M. S., Dang, M., Grimm, R., Wigginton, R., **Espinosa, J. E.**, Diguilio, L., Hunt, E., Baldwin, E. E., & Harmon, L. (2014). Online Instruction Pilot Project (OIPP) final evaluation report. Santa Barbara, CA: University of California Educational Evaluation Center.

Espinosa, J. E. (2013). Condor Technology Internship Program: 2013 evaluation report. University of California, Santa Barbara: Office of Education Partnerships. http://www.oxnardcollege.edu/departments/academic/title-v-hsi-stem-grant/project-and-activity-evaluations

CONFERENCE PRESENTATIONS

Duran, R., **Espinosa, J. E.**, Rodriguez, L., Castellanos, M., Rogel, Z. (2014, October). *Empowerment and participatory evaluation approaches: strategies for successful implementation*. Presented at the annual meeting of the American Educational Research Association, Denver, CO.

Espinosa, J. E., & Yun, J. T. (2013, April). *Assessing students' behavioral, affective, and cognitive engagement in higher education settings: An initial instrument validation.* Presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Espinosa, J. E., Dang, M., & Yun, J. T. (2012, November). *Towards a new engagement framework in higher education: Comparing students' course engagement to instructors' predicted engagement outcomes.* Presented at the annual conference of the Association for the Study of Higher Education, Las Vegas, NV.

Espinosa, J. E., & Yun, J. T. (2012, October). *Student engagement: Uncovering behaviors, feelings, and thoughts in online higher education courses.* Presented at the annual conference of the American Evaluation Association, Minneapolis, MN.

Espinosa, J. E. (2012, October). Session Chair, *Distance Education and Education Technology in Higher Education Assessment.* Multi-paper session sponsored by the Assessment in Higher Education and the Distance Education & Other Technologies Topic Interest Group, American Evaluation Association, Minneapolis, MN.

ACADEMIC SERVICE

UNIVERSITY OF CALIFORNIA, SANTA BARBARA

The Gevirtz School Diversity & Equity Committee — 2011 – 2013
Educational Leadership & Organizations Representative

UC-HBCU Initiative: UCSB-FAMU Partnership — 2011 – 2013
Committee Member

SAN JOSE STATE UNIVERSITY

Undergraduate Social Work Association — 2009 – 2010
President

School of Social Work Graduation Committee — 2009 – 2010
Co-Chair

AWARDS & RECOGNITIONS

UNIVERSITY OF CALIFORNIA, SANTA BARBARA

The Gevirtz School Dissertation Fellowship	2015 – 2016
Phillip & Aida Siff Educational Foundation Scholarship	2014 – 2015
Mosaic–UCEC Innovative Evaluation Research Award	2012 – 2013
Association for the Study of Higher Education: Council on Ethnic Participation's Mentor-Protégé Program	2012 – 2013
The Gevirtz School Continuing Student Fellowship	2011 – 2012
Graduate Student Academic Advancement: Scholarly Collaborations Grant	2011 – 2012
The Gevirtz School Recruitment Fellowship	2009 – 2010

SAN JOSE STATE UNIVERSITY

President's Scholar Recipient	2009 – 2010
Bachelor of Arts in Social Work Outstanding Project Award	2009 – 2010
Educational Opportunity Program: Academic Excellence Certificate	2007 – 2010
Dean's Scholar Recipient	2007 – 2009

PROFESSIONAL AFFILIATIONS

American Education Research Association
American Evaluation Association
Association for the Study of Higher Education

TRAINING & SKILLS

Quantitative Methodology
- Introductory Statistics
- Inferential Statistics
- Linear Models
- Advanced Multivariate Statistics
- Factor Analysis
- Structural Equation Modeling
- Survey Research & Design

Software Proficiency
- SPSS
- Mplus
- Dedoose
- Express Scribe
- Microsoft Office Suite

Qualitative Methodology
- Qualitative Methods
- Qualitative Interviewing

Areas of Study
- Educational Leadership
- Organizational Change & Development
- Organizational Theories
- Education Policy
- Higher Education Policy
- Program Evaluation
- Student Affairs Administration
- Cognitive Perspectives on Achievement Motivation

ABSTRACT

Assessing the Factorial Validity, Measurement Invariance, and Latent Mean Differences of a Second-Order, Multidimensional Model of Academic and Social College Course Engagement: A Comparison Across Course Format, Ethnic Groups, and Economic Status

by

Juan Emilio Espinosa

The current study seeks to validate a second-order, multifaceted model of engagement that contains a behavioral, an emotional, and a cognitive subtype as proposed by Fredericks, Blumenfeld, and Paris' (2004), while also incorporating literature on student interactions. The second-order, 12-factor model proposed and tested for its validity partitioned engagement into the second-order constructs of academic and social engagement and examined each of the three engagement subtypes in relation to the interactions that students experience with their course material, with their classmates, and with their instructors/teaching assistants. Since the proposed model did not meet accepted standards of fit, the dataset was randomly split into two approximately equal halves and a follow-up exploratory factor analysis (EFA) was conducted on the first half of the dataset, which yielded a second-order, five-factor solution. The second-order academic engagement constructs that emerged from the EFA consisted of students' behavioral, emotional, and cognitive engagement with their course material. In addition, two first-order factors emerged

from the EFA, consisting of students' emotional and cognitive engagement with their fellow students or classmates.

These constructs and relationships were consistent with the theory that drove the original proposed model, but differed slightly in their composition and relationship with one another. After establishing this empirical model through EFA procedures, the model was cross-validated on the second-half of the randomly split dataset and examined for invariance across students enrolled in online courses and students enrolled in traditional, in-person college courses, as well students from ethnically and economically diverse backgrounds. Latent mean comparisons revealed differences in levels of academic and social engagement between these three groups of students, suggesting that students enrolled in online courses and students from African-American and Latino/a ethnicities were slightly more academically engaged than their counterparts. However, students enrolled in online courses scored much lower than students enrolled in face-to-face courses on the social engagement measures, while students from African-American and Latino/a ethnic groups scored higher on the social engagement measures than did students from Asian and Caucasian ethnicities. Interestingly, no differences emerged between groups of students from lower and higher economic backgrounds.

TABLE OF CONTENTS

TABLE OF FIGURES

CHAPTER 1.0. INTRODUCTION

Education is capable of having profound effects on a person's life, particularly for students who persist through their studies and attain a college degree or higher. A few of these direct impacts include economic benefits and job security. The average rate of unemployment in the United States (U.S.) in 2015 was a mere 2.8% among individuals 25 years and older with a bachelor's degree; however, among individuals with less than a high-school diploma, the average unemployment rate was nearly three times (2.86) as high at 8.0% (U.S. Department of Labor, Bureau of Labor Statistics [BLS], 2016). Similarly, the median weekly earnings for individuals with a bachelor's degree were approximately 2.3 times higher than they were for individuals with less than a high school diploma. Individuals with a college degree had median weekly earnings of $1,137, whereas, individuals with less than a high-school diploma had median weekly earnings of $493 (BLS, 2016). In addition to the economic benefits associated with higher levels of education, people with a college education have been found to score higher on indicators measuring quality of life including happiness, life satisfaction, and overall health (Pascarella & Terenzini, 2005). The benefits associated with higher educational attainment may be the reason higher educational institutions continually attract large numbers of students.

According to the Digest of Education Statistics 2013, an annual report published by the National Center for Education Statistics (NCES), a whopping 20.6 million students were enrolled in a U.S. college or university in fall 2012 (Snyder & Dillow, 2015). Of these 20.6 million students, roughly 86.0% or 17.7 million students were undergraduate students, which represented a 24 percent increase in undergraduate college enrollment since 2002 when only about 14.3 million undergraduate students were enrolled in college (Snyder & Dillow, 2015). While the total number of students enrolled in college has declined 2.0% since 2010, college enrollment is expected to further rise throughout the next decade. The NCES projects college enrollment will reach new records and increase an additional 15 percent between fall 2015 and fall 2023 (Snyder & Dillow, 2015). In addition to overall enrollment, college enrollment among students from certain racial groups is vastly higher today than it was several decades ago.

College enrollment among students from all ethnic groups and economic backgrounds has steadily increased over the past several decades, but students from African-American and Latino/a racial groups are pursuing postsecondary education at much higher rates than they were several decades ago. Latino/a students enrolled in any undergraduate degree-granting institution nearly quadrupled between 1990 and 2013. In 1990, only about 700,000 Latino/a students were enrolled in an institution of higher education. This number increased approximately 4.14 times to 2.9 million students in 2013 (Kena, et al., 2015). Similarly, the number of African-American students enrolled in any undergraduate degree-granting institution more than doubled between 1990 and 2013, from 1.1 million to 2.5 million students (Kena, et al., 2015). Students from Caucasian racial backgrounds have traditionally constituted the largest group of college-going students, and this trend continues today. As of 2013, college enrollment among Caucasian students (9.9 million) was more than three times (3.4) higher the number of Latino/a students enrolled in college, and Latino/a students are the second largest group of students currently attending college. Students from Asian ethnicities are at the lower end of the college enrollment spectrum, as approximately 1.0 million from Asian ethnicities were enrolled in a U.S. college or university during 2013 academic year (Kena et al., 2015). Despite the variation in college enrolment among different groups of students, colleges are responsible for serving a significant body of students.

1.1. The Rise of Online Education

Online education is one form of instruction that may help colleges meet enrollment demands, since online education minimizes the physical capacity needed to educate students. Since 2003, the annual growth rate of students enrolling in at least one online course exceeded the overall college enrollment growth rate (Allen & Seaman, 2015). While enrollment data indicates that online course enrollment rates have dwindled over the past several years (Allen & Seaman, 2015), there were still approximately 5.5 million students who decided to enroll in some form of distance or online education course during fall 2013 (U.S. Department of Education, National Center for Education Statistics, Integrated Postsecondary Education Data System [IPEDS], Dec. 2014). Furthermore, more than 70 percent of all public, degree-granting institutions of higher education offered online courses, and more than 95 percent of colleges with more than 5,000 students offered some sort of online instruction (Allen & Seaman, 2015). In California, the rapid growth of online education is driven by the desire to increase student access and enrollment, while also reducing costs for the state (Johnson, Mejia, & Cook, 2015). It seems likely that higher education institutions in other states are embracing online education for similar reasons.

The academic experiences and course outcomes of students enrolled in online classes has not been well documented among the research community. Even fewer studies have examined the experiences of underrepresented minority students and students from low-income backgrounds in these types of courses. Despite the limited literature that has been conducted on students enrolled in online courses, students enrolled in online classes have been found to have lower retention and academic success rates than students enrolled in traditional, face-to-face courses (Carr, 2000; as cited in Jaggars & Bailey, 2010). Although the experiences of students from different racial groups and economic backgrounds in online course has not been well studied, students from Latino/a and African-American racial groups as well as students from lower economic backgrounds have historically had lower college success and completion rates than their racial and economic counterparts (Carter, 2006; Pascarella & Terenzini, 2005). Unfortunately, the academic challenges that these groups of students have encountered remains an issue within our educational system today. The academic issues encountered by these groups of students has been studied for decades; however, more research must be conducted to determine the applicability of critical student success and retention theories in online courses settings.

1.2. The Complexity of Student Engagement

There are a range of approaches and methods that may be helpful in supporting students' academic success. Student engagement is one approach and aspect of education that has received a considerable amount of attention among the research community across all levels of education. This examination appears to be motivated by the belief that student engagement is capable of positively predicting students' academic achievement and retention (Fredericks et al., 2004; Kuh, Kinizie, Buckley, Bridges, & Hayek, 2006; Pascarella, 2010). While student engagement has been extensively investigated, there has been much variation in approaches used to study the concept. Researchers studying student engagement at the primary and secondary education level often measure student engagement through behavioral, emotional, and/or cognitive components and examine the extent to which students engage with their institution, in their course, and during specific learning tasks (Fredericks, Blumenfeld, & Paris, 2004). Researchers at the post-secondary education level often assess student engagement based on concepts found in the *Seven Principles of Good Practice in Undergraduate Education* proposed by Chickering and Gamson (1987) and Astin's (1984) *Theory of Student Involvement*, drawing particular attention to the interplay between students, their classrooms, and the larger school context (Kuh, 2001; Robinson & Hullinger, 2008). Although research at the primary and secondary education level have examined student engagement both at the classroom and institutional level (Finn, 1989; Goodenow, 1992; Marks, 2000; Newmann, Wehlage, & Lamborn, 1992; Skinner & Belmont, 1993; Skinner, Kinderman, & Furrer, 2009), fewer studies at the post-secondary education level examine student engagement within the context of a specific course; instead, they focus on examining students' engagement with their broader educational institution (Kuh, 2001; Pascarella, Seifert, & Blaich, 2010).

Reviewing online or distance education literature revealed that student engagement was also an area of interest and study among researchers in this arena. Unlike research conducted on student engagement at the K-12 and post-secondary education levels, a common framework for the assessment of student engagement in online course settings was not found. The National Survey of Student Engagement (NSSE) is an instrument that has been widely used by researchers and institutions to assess student engagement in college settings; however, this tool must be modified or adapted in order for it to maintain relevance to students enrolled in online courses since it focuses largely on student institutional engagement and not engagement at the course level (Chen, Lambert, & Guidry, 2010; Dixson, 2010; Kuh et al., 2006; Pascarella, Seifert, & Blaich, 2010; Robinson & Hullinger, 2008). Other researchers have developed their own measures to study engagement (Dixson, 2010). The range of methods applied when examining student engagement adds to the complexity of the student engagement construct.

1.3. Theoretical Foundations of the Current Dissertation

The growth of online education demands that important theories and frameworks that have traditionally be devised for students enrolled in brick and mortar settings be tested for their relevance and applicability to online course settings. The current study attempts to achieve this goal by proposing, testing, and validating a student engagement measurement model and examining the extent to which the model is applicable to students enrolled in online and face-to-face college courses, students from differing racial groups, and students from higher and lower economic backgrounds. When developing the engagement model, which I have named the Engagement Measurement Model of Students' College Course Success, I reviewed the vast literature on student engagement and distance learning across all levels of education. For this study, I adapt the tripartite, multidimensional engagement framework proposed in Fredericks and colleagues' (2004) seminal review of primary and secondary literature on student engagement. Similar to Fredericks and colleagues' framework, engagement in this study will be tested as a multidimensional construct containing three engagement subtypes—a behavioral engagement subtype, an emotional engagement subtype, and a cognitive engagement subtype. To ensure that the model was applicable to students enrolled in online and face-to-face college courses, I reviewed distance education literature and made appropriate modifications to the model. Towards this end, the engagement model that I will test in the current study examines the extent to which students engage behaviorally, emotionally, and cognitive with their course content or material, with their classmates, and with their instructors/teaching assistants (TAs).

Meaningful interactions between course members are critical to the development of engaging course environments at all levels and in all forms of education; however, these aspects are particularly important when creating engaging online course environments, since students in online classes do not share the same physical space with other course members. Students typically encounter three types of interactions in college courses, which are referred to as student-content, student-student, and student-instructor interactions in the distance education literature (Bernard et al., 2009). This study rests on the assumption that students are capable of engaging behaviorally, emotionally, and cognitively with their material, with their classmates, and with their instructors/teaching assistants (TAs). Academic forms of engagement in the current study pertain to students' behavioral, emotional, and cognitive engagement with their course content or material, while social forms of engagement refer to the three engagement subtypes that students experience with their classmates and with their instructors/TAs. Together, the engagement model components represent nine of the 12 latent constructs that I will test through a confirmatory factor analysis (CFA) to determine suitability of the model.

1.4. Key Definitions of Engagement Model Components

The review of engagement literature revealed that researchers have used many methods and approaches when assessing these three-engagement subtypes. While a more thorough review and explanation of the approaches that researchers have used to measure these forms of engagement is provided in the next chapter, the following definitions were applied to each engagement subtype included in the current model. These definitions were consulted when creating survey items to represent students' behavioral, emotional, and cognitive engagement. *Students' behavioral engagement with their course material* refers to students' involvement or participation in class activities and requirements; *students' emotional engagement with the course material* pertains to students' affective reactions towards the class material and activities; and *students' cognitive engagement with the course material* refers to the cognitive and metacognitive processes that students utilize in order to better understand the course material and concepts. Similar definitions were applied to students' engagement with their classmates and instructors/TAs. *Students' behavioral engagement with their classmates and instructors/TAs* investigates the extent to which students interact with either of these course actors; *students' emotional engagement with their classmates and instructors* refer to students' attitudes and feelings towards the interactions they experience with these individuals throughout their course; and *students' cognitive engagement with their classmates and instructors/TAs* assesses whether students' interactions with either of these course members increased their understanding of the course material and concepts.

The engagement-centered measurement model that I propose for this study also contains two latent constructs that I believe represent specific types of pedagogy and another construct that I believe reflects students' course satisfaction. The first pedagogical construct relates to the organization and structure of the course, which I have labeled effective instructional or course design. The second pedagogical construct will assess instructors' facilitation of interactive or collaborative learning activities, which I have labeled collaborative learning. Students' course satisfaction pertains to students' attitudes and feelings about their experience throughout the entire course. These additional constructs will be examined simultaneously with the nine previously mentioned engagement subtypes to determine if the data adequately represents the proposed engagement measurement model.

1.5. Research Aims, Questions, & Hypotheses

The current study seeks to validate a multidimensional, engagement-centered measurement model through covariance analyses. Based on an extensive review of educational literature, I propose a comprehensive approach for examining student engagement in college courses. Fredericks, Blumenfeld and Paris' (2004) framework served as the basis for the current model, but the model that I will propose, test, and attempt to validate examines student engagement with greater precision by examining student engagement throughout the types of interactions that students experience with their course content/material, their classmates, and their instructors. In addition, I also hope to validate the additional constructs that relate to pedagogy and course satisfaction, which would provide the foundation for examining predictors and outcomes of specific forms of engagement in future studies. Thus, the first goal this study is to propose, test, and validate a multidimensional engagement measurement through factor analytic approaches, which contains nine constructs that pertain to students' behavioral, emotional, and cognitive engagement with the course material and with various course actors, two constructs that pertain to pedagogical or instructional methods, and one construct that pertains to students overall course satisfaction.

Another central goal of the current study is to determine the extent to which the engagement model is applicable to students enrolled in different college course settings and to students from differing ethnic and economic backgrounds. Prior to proceeding with this aspect of the study, I must first establish an engagement measurement model that is supported both empirically and statistically. After establishing an engagement-based measurement model, I will continue my analysis by examining the model of measurement invariance in a CFA framework. More specifically, I will conduct a multiple-group confirmatory factor analysis (MG CFA) across the following groups of students to determine the extent to which the model is applicable to these students in the study:

1. Students enrolled in online courses and students enrolled in face-to-face courses;
2. Students from ethnicities who have historically performed well academically (i.e., Caucasian or White and Asian students) and students from ethnicities who have not performed as well as their racial counterparts (i.e., African-American or Black and Latino/a students); and
3. Students classified by university data as being low income and students not classified by university data as being low income.

I developed the following series of research questions to guide the work that I performed for this dissertation and to address the goals of this study.

1. Does covariance analysis validate the existence of a 12-factor engagement measurement model, which contains nine course engagement subtypes, two constructs that relate to pedagogy, and one construct that pertains to students' course satisfaction?
    a. If the proposed model is not validated through a confirmatory factor analysis, does an exploratory factor analysis support the existence of an alternative second- engagement measurement model?
2. Does covariance analysis support the existence of a multidimensional engagement measurement model that contains behavioral, emotional, and/or cognitive engagement components?
    a. Do the engagement components exist between students and their course material, students and their classmates, and students and their instructors/TAs?
3. Do tests of measurement invariance, conducted through a variance-covariance multiple-group (MG) CFA framework, suggest that the following groups of students are interpreting the final engagement measurement model similarly?
    a. Students enrolled in online courses and students enrolled in face-to-face courses;
    b. Students from ethnicities who have historically performed well academically (i.e., Caucasian or White and Asian students) and students from ethnicities who have not performed as well as their racial counterparts (i.e., African-American or Black and Latino/a students); and

c. Students classified by university data as being low income and students not classified by university data as being low income.

I hypothesize that the engagement measurement model that I propose for this study will, indeed, meet accepted standards of fit. Furthermore, I also hypothesize the model will prove to be a second-order measurement model in which the nine first-order engagement constructs will be represented by the second-order constructs of academic engagement and social engagement. More specifically, I anticipate that the second-order academic engagement factor will be represented by students' behavioral, emotional, and cognitive engagement with their course content. I also hypothesize that the second-order social engagement factor will be represented by six first-order constructs, consisting of student-to-student and student-to-instructor/TA behavioral, emotional, and cognitive engagement. My last hypothesis for this study is that the engagement measurement model that transpires will function invariantly across all groups of students that I identified for this dissertation, suggesting that the model may have the potential to identify sources and forms of engagement across diverse college course settings and groups of students. Ideally, this information will be used by instructors, course designers, and/or college personnel to gauge students' levels of academic and social engagement and determine whether these levels coincide with the instructors' intentions for student engagement throughout a course.

1.6. Study Implications

The current study will contribute to the literature on student engagement by providing a comprehensive approach for examining engagement in multiple settings and across multiple groups of students at the course level that can provide information for instructors and online course designers about how students are interacting with their courses. The specificity of the proposed model will allow sources of engagement to be identified. Most existing measures of engagement group together specific sources of engagement into single measures, which prevents educators from identifying the extent to which students are engaging with their course material or with other course members. The proposed model seeks to address this problem with the measurement of engagement by distinguishing academic from social forms of engagement. An instrument with this level of information will allow instructors to assess specific sources of engagement and modify or adjust specific activities as needed. A model that is found to function invariantly will yield additional contributions to our understanding of student engagement since it is likely to be valid across multiple contexts and situations.

The current study is conducted at a time in which technology is rapidly evolving. The surge in online course offerings among institutions of higher education draws an immediate need to determine whether fundamental theories and frameworks that have been primarily examined in traditional, brick and mortar classroom settings are suitable to online course environments. The current study will begin to shed light onto whether Fredericks and colleagues' (2004) conceptualization of engagement is applicable to online settings. Validation of the current model will provide support for further studying engagement as a multidimensional construct in online college course settings to uncover the relationship between specific types of engagement and more distal student outcomes. The tests of invariance will also shed light on the extent to which the model functions among students from different ethnic and economic backgrounds, which may ultimately be used characterize the course engagement among these students and identify specific forms of engagement that are correlated with the course satisfaction and success of these students.

CHAPTER 2.0. LITERATURE REVIEW

A review of the literature on student engagement revealed that there has been a range of methods and measures applied when assessing student engagement. These approaches have differed among researchers studying student engagement at the primary and secondary education level and researchers studying student engagement at the post-secondary education level. As synthesized in a review of primary and secondary educational research by Fredericks and colleagues (2004), behavioral, emotional, and/or cognitive engagement are frequently used to assess student engagement. The current study adapts this model for examining student engagement and applies it to online and traditional, face-to-face college courses. The following sections summarizes key literature that contributed to the development of the engagement measurement model that I propose and test in this study. I begin this review by detailing the engagement subtypes proposed by Fredericks, Blumenfeld, and Paris (2004) and the variations in which these forms have been studied and measured. I continue by explaining methods used to measure student engagement in post-secondary education settings as well as online higher education courses. I also review key literature pertaining to pedagogy in online or distance education settings, which also helped shape the model proposed in this study. I conclude this review by summarizing the goals of the current study, illustrating the engagement measurement model that I will test, and detailing the operationalization of components in the engagement model.

2.1. Behavioral Engagement

Behavioral engagement has been recognized as a component of student engagement in most the primary and secondary educational literature reviewed;

however, researchers have differed in the manner in which they have measured the construct (Downer, Rimm-Kaufman, & Pianta, 2007; Fredericks et al; 2004). When measuring behavioral engagement, some researchers have focused on the psychological components and applied indicators such as effort, attention, and persistence (Fredericks et al., 2004; Marks, 2000; Skinner & Belmont, 1993; Wang, Willet, & Eccles, 2011). Studies that have focused on psychological components often explore the relationship between student engagement and motivation (Goodenow, 1992; Meece, Blumenfield, & Hoyle, 1988; Ryan & Patrick, 2001). There are differing opinions among educational experts on whether motivation and student engagement are distinct (Appleton, Christenson, Kim, & Reschly, 2006). Those who view motivation and engagement as being distinct do not deny that there may be a relationship between the constructs; however, they assert that motivation is not necessary for a student to be engaged (Appleton, Christenson, & Furlong, 2006).

Some researchers have applied psychological indicators when assessing behavioral engagement, while others have focused specifically on observable behaviors such as class participation, coursework completion, course attendance, and classroom conduct (Appleton, et al., 2006; Finn & Voelkl, 1993). Focusing specifically on observable behaviors may be one method to draw the distinction between student engagement and motivation. Once these two constructs are disentangled, associations between different motivation assessments and levels of student engagement may be explored. In addition, focusing specifically on observable behaviors is easier to measure than psychological processes. While there has been some variation in the measurement of this engagement subtype, behavioral engagement or simply student participation is an element critical to student success, and some have argued that student success is entirely dependent on their level of academic involvement (Astin, 1984). It may be equally important for students to develop and maintain positive perceptions, attitudes, and feelings, about their coursework, classmates, instructors, and institution.

2.2. Emotional Engagement

The second component of the multidimensional model proposed by Fredericks and colleagues (2004) is emotional engagement. Emotional engagement reflects students' affective responses to various aspects of their class. Researchers often merge students attitudes and feelings towards various aspects of school into a single construct when measuring emotional engagement. Jimmerson, Campos, and Grief (2003) wrote an article in an attempt to clarify the construct of school engagement and measures often associated with the construct. They defined emotional or affective engagement as students' feelings about their school, teachers, and/or peers. They explained that school bonding is an indicator that is often used to assess students' emotional engagement; however, they noted that school bonding and related terms—such as belonging, school community, school membership, motivation, and school attachment—are not always defined. School bonding and these similar terms pertain to students' connection to their educational institution. Although there are slight differences between school bonding and related terms, they share similarities in that they explore students' feelings of inclusion and connectedness with their classmates, instructors, and/or institution.

The construct of emotional engagement has been examined between students and different school members, such as their peers and instructors, and within different contexts, such classrooms and institutions. These approaches to measuring emotional engagement suggest that environmental factors play a role in determining students' level of engagement. Specifically, students' social context plays a significant role in their learning and development, (Wentzel, 2004). Students are constantly interacting with their peers, their teachers, and their school personnel; the quality of these interactions and relationships influence students' perception of their fit within their school and class environment (Tinto, 1975; 1988). These perceptions likely influence their overall levels of school satisfaction, which, consequently, influence their levels of academic engagement and decision to persist through their studies (Bean & Eaton, 2001). Conversely, students who do not fit into their school environment nor get along with their peers and/or instructors may demonstrate lower levels of engagement and suffer academically by submitting low-quality work, disengaging, and/or dropping out (Rumberger, 2001; Tinto, 1975). Hence, methods to elicit positive emotional stances that students feel towards their course content, classmates, and instructors should improve students' academic experience and success.

Measures used to assess emotional engagement often combine students' attitudes and feeling towards their classmates, instructors, and broader institution. Educational administrators, instructors, and practitioners would benefit from having a more detailed understanding of students' emotional stances towards each of these aspects of school. As such, the model in this paper examines students' engagement with these areas independently. This is particularly important since students may develop connections and emotional ties with their classmates but not their instructors and vice versa. While the emotional connections students develop with their instructors and classmates may deepen students' connection with their school, it also important for these interactions to help students comprehend the course material being learned. As such, the model tested in the current study also includes assessments of students' cognitive engagement with their course content, and examines the influence of social interactions on students' cognitive comprehension.

2.3 Cognitive Engagement

The conceptualization of cognitive engagement in primary and secondary education varies depending on the field of research. According to Fredericks and colleagues (2004), cognitive engagement stems from literature on learning and instruction as well as achievement motivation. It has been argued that students who are cognitively engaged are strategic and self-regulating (Corno & Mandinach, 1983; Zimmerman, 1990), psychologically invested in their learning, and willing to exert additional effort and seek challenging learning situations (Fredericks et al., 2004). Regardless of the indicators appointed to measure the construct, cognitive engagement is difficult to assess because indicators of cognitive engagement are not directly observable (Appleton, et al., 2006).

Corno and Mandinach (1983) examined cognitive engagement in classroom settings while also taking into account students' motivation. Corno and Mandinach posited that students are continually attempting to interpret the interactions that occur between themselves and their classroom environment, which influences the amount of effort that they expend towards their academic work. Thus, indicating a relationship between behavioral and cognitive engagement. They further claim that self-regulated learning is the highest form of cognitive engagement. Self-regulated learning is an attempt for a student to deepen their understanding of a particular area, while assessing and enhancing this understanding. Similarly, Weinstein and colleagues (2011) posited that self-regulated learning occurs when students apply metacognitive strategies to assess their understanding of information, monitor and regulate effective and efficient learning strategies, focus their attention, and maintain concentration. There seems to be a consensus that self-regulated learning contains three components: the metacognitive strategies that students use to plan, set goals, organize, and self-evaluate personal knowledge prior to engaging in a learning task; the manner in which students control and manage their effort during the learning task; and the cognitive strategies that students use to process and understand material after their involvement in a learning task (Clearly and Chen, 2010; Pintrich & De Groot, 1990; Zimmerman, 1990). Carefully examining definitions commonly applied to this engagement subtype suggests that cognitive engagement is largely believed to be an individual process.

While the definitions applied to cognitive engagement make the construct appear to be an independent or individual process, environmental and social factors on students' knowledge acquisition has been examined by social cognitive theorists (Bandura, 2002; Zimmerman, 1995). Future research should explore the effect of student interactions on students' cognitive engagement. Students are continually interacting and working with their peers and teachers; these interactions have the potential to enhance students' learning and development. The model that will be presented later in this paper incorporates social dimensions to students' cognitive engagement by examining whether interactions increase students' cognitive engagement and overall course content comprehension.

Behavioral, emotional, and cognitive engagement has primarily been examined at the primary and secondary education level; however, these areas have implications for higher-education settings (Lester, 2013). Most research on student engagement in higher education has focused on findings from the National Survey of Student Engagement (NSSE), which is a survey that has been adopted by many colleges to assess student engagement. The following sections further explain the NSSE and approaches that have typically been taken to assess student engagement at the post-secondary education level.

2.4. Student Engagement in Higher Education

Student engagement in higher education has been examined differently from student engagement at the primary and secondary education level; however, they do share some similarities. Engagement in college is typically measured by assessing the effort students exert towards their curricular activities. Assessments have also been developed to determine effective educational practices. Specifically, engagement has been defined as "the time and effort students devote to activities that are empirically linked to desired outcomes of college and the ability for institutions to promote students' participation in these activities" (Kuh, 2009, p. 683). The definition and assessment of student engagement is rooted in Astin's (1984*) Theory of Student Involvement* as well as Chickering and Gamson's (1987) *Seven Principles for Good Practice in Undergraduate Education*.

A significant number of colleges have administered the NSSE and utilized data from this instrument to gauge student engagement. The following five benchmarks represent indicators that are used in the NSSE to assess student engagement. (1) level of academic challenge, which focuses on the academic effort that students place towards their studies and the educational expectations institutions set for their students; (2) active and collaborative learning, which is based on the assumption that students learn most when they are involved in their academic studies and are forced to apply their learning to different settings; (3) enriching educational experiences, which is based on the notions that classroom learning should be complemented with other learning opportunities, students should experience diversity, technology should facilitate learning and promote collaboration, and students should apply their knowledge through internships and other related activities; (4) student-faculty interactions, which centers on the belief that these types of interactions allow students to examine ways that experts think about and solve practical problems; and (5) supportive campus environment, which is based on the assumption that student performance is optimized when institutions are committed to student success and devoted to providing students with positive working and social relations.

A closer examination of these five pillars or benchmarks elicits one main similarity between student engagement measures at the primary and secondary level and measures at the postsecondary education level: across both levels of education, the time and effort students expend towards their academic studies has been defined as a component of student engagement. This suggests that all student engagement assessments should measure student participation or involvement. The other NSSE benchmarks differ from measures that are typically applied at the primary and secondary education level. The NSSE framework explores students' interactions to a greater extent than does the framework proposed by Fredericks and colleagues (2004). Conversely, none of the NSSE assessments explore students' emotional responses to various aspects of school such as their peers, instructors, or institution.

Merging methods used to assess engagement across these levels of education would provide a more thorough understanding of students' engagement experience.

Most of the benchmarks incorporated into NSSE's assessment of student engagement examine institutional practices that are believed to promote student engagement. While findings from the assessment of these benchmarks may provide useful information for institutions, it is questionable to assert that these benchmarks reflect the totality of student engagement. Research has suggested that the five pillars including in the NSSE are positively correlated with successful student outcomes, including academic performance and persistence (Kuh et al., 2006; Pascarella, et al, 2010). Many aspects of these benchmarks have previously been examined and studied; however, they were not originally framed as student engagement. For example, Pascarella (1980) created a model explaining practices that promote successful student outcomes that focused on the informal interactions between students and their instructors and peers; other researchers have examined active and collaborative classroom learning (Faust & Paulson, 1998). These five benchmarks or pillars seem to reflect a number of practices that predict successful student outcomes. It appears that these practices were strategically selected, because they have been found to be positively associated with successful student outcomes. While findings from the NSSE should be helpful for institutions, the definition that has been applied to engagement at the postsecondary education level seems to be more of a fusion of effective educational practices and not student engagement. More specifically, the pillars for the NSSE do not provide data that would be directly actionable by classroom instructors to improve their own courses, or answer questions about types of courses, such as online courses. This means that the most important national survey on higher educational engagement does not have measures that can answer questions for instructors at the course level. Clearly this is an issue that needs to be addressed, particularly with the growth in popularity in online courses and the lack of understanding about how student can successfully engage with this relatively new medium. In addition to this problem of content in the NSSE, there has been some additional debate on the validity of the NSSE.

There is some evidence that the NSSE has high predictive validity on important student outcomes (Pascarella et al., 2010), but recent research has questioned the validity and reliability of the NSSE at the institutional level. Campbell and Cabrera (2011) published an article claiming that many validations of the NSSE have not focused on advanced statistical techniques to determine the number of constructs in the NSSE. They further state that the NSSE instrument would benefit from having sound statistical support through techniques such as confirmatory factor analysis or item response theory. Porter (2011) conducted a statistical analysis of the NSSE and claimed that the survey had a number of validity issues. To justify this claim, he argued that the survey was guided largely by empirical data as opposed to theoretical data; benchmarks of engagement have not been replicated by other researchers; and measures of reliability fail to meet statistical standards. A survey that is not valid or reliable would call into question a number of findings that have been extracted from its data. The work that I am engaging in here will address both of these concerns. I am grounding my engagement measure in student behaviors at the course and not the institutional level, and I will be using standard psychometric approaches to survey validation which will address both concerns with the NSSE.

Prior to discussing an alternative framework for assessing student engagement in online settings, an understanding of approaches that have been taken to assess student engagement in online courses should be understood. Most college institutions offer some form of online education; however, student retention has suffered in online courses, as retention rates have typically been lower in online courses than traditional, in-person courses (Angelino, Williams, & Natvig, 2007; Carr, 2000). Enhancing students' engagement may be one method to address this issue. The next section summarizes some of the literature on student engagement in online classes.

2.5 Student Engagement in Online Courses

There is quite a bit of variation in the research approaches to assess student engagement in online courses. While research on student engagement at the primary and secondary education level also varied, most studies could be characterized by the three-part engagement model proposed by Fredericks, Blumenfeld, and Paris (2004). Similarly, most studies conducted at the post-secondary education level on student engagement relied on the NSSE and related benchmarks. There was slightly more variation in approaches used to assess student engagement in online courses. Some researchers developed their own scales due to the lack of available measurements of student engagement in online classes (Dixson, 2010), while other researchers used components of the NSSE to serve as the foundation of for their engagement assessments (Chen, Lambert, & Guidry, 2010; Robinson & Hullinger, 2008).

Robinson and Hullinger (2008) conducted a study that utilized some of the benchmarks included in the NSSE's framework and applied it to online courses. Since the NSSE incorporates a number of measures that assess the influence of effective institutional practices on student engagement, Robinson and Hullinger (2008) modified the NSSE instrument in order to apply it to online course settings. While they incorporated most of the NSSE factors, including level of academic challenge, students' interactions with faculty members, and enriching educational experience, they omitted questions that assessed whether students were provided with a supportive campus environment. Robinson and Hullinger decided to exclude these items, because they were not relevant to online classes. The need to adapt engagement measures demonstrates the need for alternative approaches for examining engagement in higher education and online settings. Researchers would benefit from having a model that can be used at all levels of education and styles of courses to measure student engagement.

Dixson (2010) developed scale to measure student engagement in online courses and provided the following justification for developing the scale: "Because there was no scale to measure online student engagement, the first stage of the project was to develop a measure of student engagement in online classes" (p. 3). To develop their online engagement scale, Dixson consulted two student engagement instruments as well as an instrument that measured students' interactions in online courses. Students who had opportunities to interact with their peers and their instructors were more engaged than students who did not have these opportunities and felt stronger emotional connections with their peers and instructors (Dixson, 2010). Ensuring students interact with their peers and instructors are critical in providing students with quality learning experiences in online classes. This study highlights the possibility of incorporating measures that examine students' interactions and emotional stances towards their peers and instructors when examining student engagement in online classes.

Many of the articles reviewed on student engagement in online classes have noted the importance of students' interactions with their classmates and instructors (Dixson, 2010; Robinson & Hullinger, 2008). These studies have focused on the positive impacts these interactions have on students' emotional stance towards their peers and instructors and the impact they have on students' class participation and involvement. Since there does not appear to be a standard approach to assess student engagement in online courses, reviewing online and distance education literature should provide greater insight on effective ways to assess student engagement in these types of classes.

2.6. Student Interactions

Bernard and colleagues (2009) conducted a meta-analysis to examine the different types of interactions that occur in online classes. Three forms of interactions frequently transpired in the studies they reviewed, which are named student-to-content, student-to-student, and student-to-teacher interactions. Student-to-content interactions occur when students interact with the material being taught in the course; student-to-student interactions occur when students work with their peers in small groups or one-on-one, and these interactions may be synchronous or asynchronous; student-to-instructor interactions occur when students interact with their instructors, which typically provide students with emotional or motivational support. Student-teacher interactions may also be synchronous or asynchronous; during these interactions

A major criticism of online courses is that they fail to provide students with sufficient means for student-to-student and student-to-instructor interactions (Bullen, 1998). Fortunately, technological advances allow real-time interactions to occur and permit instructors to replicate interactions that were once only possible when individuals shared the same physical space. Despite these advances, some still believe interactions in online environments do not provide students with the same quality of interactions that occur in face-to-face settings (Sanders, 2006). Since online courses will continue to be offered, it is not helpful to explore whether online interactions are as effective as in-person interactions; instead, methods to develop meaningful interactions in online environments must be determined. Simply providing student with avenues to interact is unlikely to lead to meaningful interactions. Instead, instructors must develop a culture that promotes and encourages students to interact with other class members.

2.7. Pedagogy & Student Engagement

Creating social presence is one method instructors may implement to encourage students to interact with their instructors and peers. Tu and McIssac (2002) provided the following definition for social presence: "Social presence is a measure of the feeling of community that a learner experiences in an online environment" (p. 131). It is believed that by developing social presence, in any type of class, students will feel greater levels of comfort between their peers and instructors, which should enhance levels of comfort among course members, increase the frequency of interactions, lead to more information sharing between class members, and improve educational outcomes (Aragon, 2003). It was noted that students and instructors alike play an important role in developing social presence.

The influence that instructors have in shaping student involvement and emotional connections resulted in me attempting to validate pedagogical constructs in the proposed engagement model.

2.8. Pedagogy & Collaborative Cognitive Engagement.

The interactions students experience with their peers and instructors are capable of providing students with a range of benefits. The benefits associated with these interactions include increasing motivation, promoting active learning, enhancing critical thinking, and improving learning outcomes (Baker, 2010; Gokhale, 1995; Lundberg & Schreiner, 2004; Johnson & Johnson, 1986). While a number of researchers have found collaborative learning, academic-based social interactions, and community to provide students with a range of benefits, others have found these interactions to be correlated with information and cognitive overload (LaPointe & Gunawardena, 2004). This cognitive overload may result in students applying surface approaches to learning instead of deep or higher-order cognitive approaches. These findings support some researchers' position that merely placing students in groups will not enhance their learning (Garrison, Anderson, & Archer, 2001; Garrison & Cleveland-Innes, 2005; Kreijns, Kirschner, & Jochems, 2003). Instead, interactions must be carefully designed to enhance students' emotional connection, while also increasing their cognitive understanding of the material.

Collaborative learning has the ability to promote deeper learning, critical thinking, collective understanding, and long-term comprehension of the information and concepts being conveyed; however, in order for interactions to affect student learning and development positively, interactions must be structured and designed to shape students' thinking and thought processes in a critical and reflective manner

(Garrison, Anderson, & Archer, 2000; Garrison & Cleveland-Innes, 2005; Kreijns et al., 2003). Simply transmitting information seems unlikely to produce a body of students who are cognitively engaged and invested in their course. Students are more likely to become cognitively engaged if their interactions with their course material, peers, and instructors are structured in ways that allow them to reflect on the course material and explore and analyze ideas (as cited in Garrison & Cleveland-Innes, 2005). Instructors' ability to create engaging learning environments and promote meaningful interactions influenced the decision my decision to incorporate pedagogical constructs that assess course structure and the facilitation of interactions distinctly, particularly since it has long been accepted among social learning theorists that social interactions are required to stimulate advanced levels of cognitive functioning, thought processes, and intelligence (as cited in Dai & Sternberg, 2004; Siegler & Alibali, 2005; Shaffer, 2005). Instructional approaches that influence students' course engagement should illuminate critical information capable of benefiting a range of students.

2.9. Overview of the Current Study

The engagement model to be tested in the current study for construct validity and its applicability to students enrolled in various course settings and to students from differing ethnic and economic backgrounds was guided by literature on student interactions, online education, and student engagement. The engagement measurement model that I will test through confirmatory factor analytic methods is illustrated in Figures 1 and 2. These figures depict the 12 latent constructs and the number of indicators that I believe will represent each of these latent constructs. Fredericks, Blumenfeld, & Paris' (2004) model served as the foundation for this study, because it was agreed that student engagement could be characterized by behavioral, emotional, and cognitive components. When developing this engagement framework, a comprehensive understanding of students' experiences was desired. As such, the framework also draws from literature on student interactions.

As previously detailed, the three main forms of interactions that occur in online courses are student-to-content, student-to-student, and student-to-instructor (Bernard et al., 2009). The current framework relies on the assumption that students may engage behaviorally, emotionally, and cognitively with their course content, classmates, and instructors or teaching assistants (TAs). Survey items were developed to assess each of these engagement areas. In order to assess the construct validity of the proposed measurement model, a confirmatory factor analysis will be conducted on the measurement model illustrated in Figures 1 and Figures 2. Alternative factor analytic methods will be employed if the current model is not found meet acceptable standards of fit. After establishing a sound measurement model, a multiple-group CFA will be conducted on the following groups of students to determine the suitability of the model to students enrolled in different course formats and from different ethnic and economic backgrounds:

1. Students enrolled in online courses and students enrolled in face-to-face courses;
2. Students from ethnicities who have historically performed well academically (i.e., Asian and Caucasian or White students) and students from ethnicities who have not performed as well as their racial counterparts (i.e., African-American or Black and Latino/a students); and

3. Students classified by university data as being low income and students not classified by university data as being low income.

Prior to explaining the methodology employed during the current study, a review of each aspect of the model is provided. Conceptualizations of each model component were largely derived from the literature recently summarized.

2.9.1. Conceptualizations of Pedagogical Model Components. Researchers studying various facets of online and face-to-face courses have explored many subtopics related to pedagogical or instructional approaches. In the present study, I attempt to validate two pedagogical constructs, because instructors play a critical role in providing students with quality learning experiences and promoting student engagement. The validation of these items will also provide researchers and practitioners with concrete evidence for further examining the influence that these forms of pedagogy have on academic and social forms of engagement. I have labeled the first pedagogical construct effective course design. This aspect of the model pertains to the structure and delivery of courses and occurs when instructors provide students with high-quality learning material, clearly articulate course requisites, and structure courses in intuitive ways. Courses that are effectively designed should be correlated with academic forms of behavioral, emotional, and cognitive engagement.

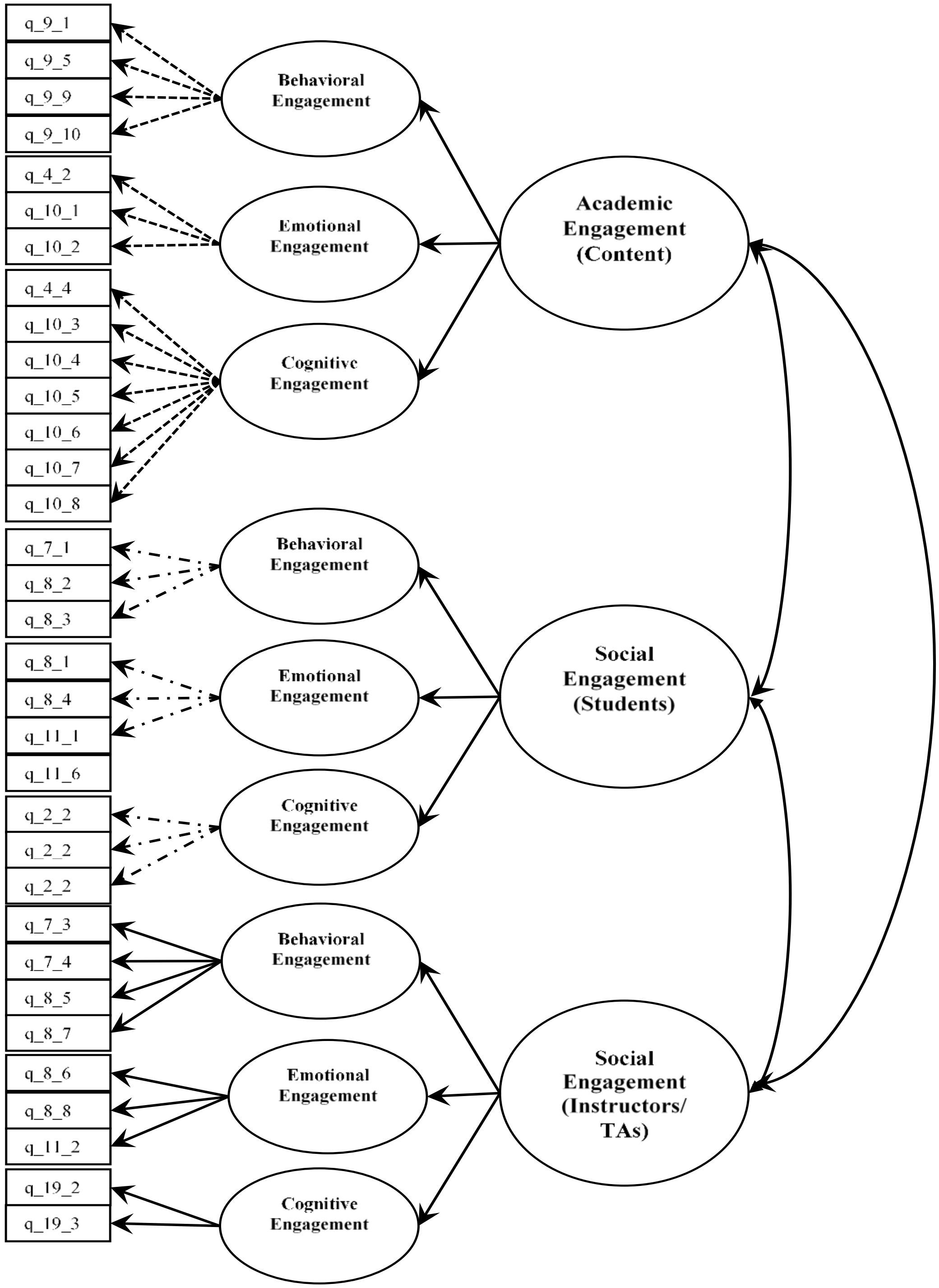

Figure 1. Academic and social engagement constructs a part of the hypothesized 12-factor CFA model of student engagement.

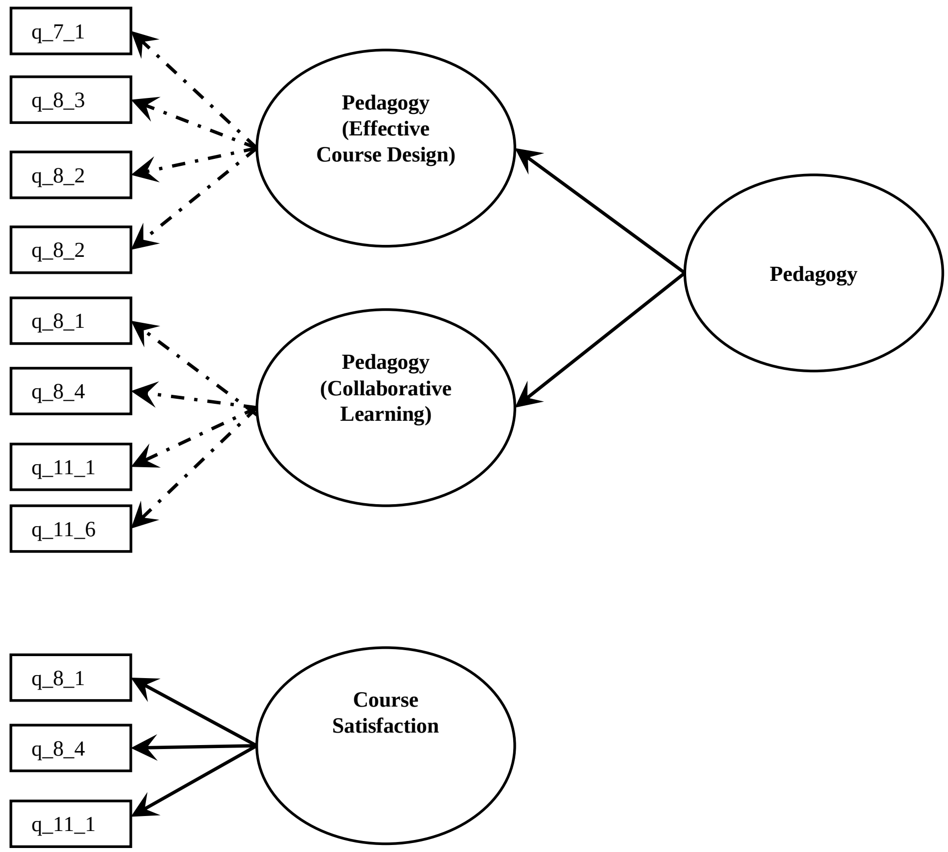

Figure 2. Pedagogy and course satisfaction constructs a part of the hypothesized 12-factor CFA model of student engagement.

The second pedagogical approach pertains to the promotion of interactive or collaborative learning activities that instructors incorporate in their course to increase student-student and student-instructor interactions, which I have termed collaborative learning. Instructors who promote interactive or collaborative learning activities should benefit by producing a body of students who are more motivated, engaged, satisfied, and successful in their course (Komarraju, Musulkin, & Bhattacharya, 2010; Kuh, et al., 2006; Pascarella & Terenzini, 1980). Thus, I believe this form of pedagogy will be positively correlated with students' behavioral, emotional, and cognitive engagement with their classmates and instructors.

As previously noted, student engagement is a complex construct that has been examined slightly differently across levels of education (Dixson, 2010; Fredericks et al., 2004; Kuh et al., 2006). While there has been much variation in the assessment of student engagement, researchers frequently find student engagement positively predicts students' academic success (Chen et al., 2010; Fredericks et al., 2004; Kuh et al., 2006; Kuh et al., 2008). This relationship motivated the decision to adapt Fredericks and colleagues (2004) engagement framework and test the modified version of their model to determine its applicability to students enrolled in online, college course environments. Since Fredericks and colleagues' three-part model has primarily been used by researchers at the primary and secondary education level, the operationalizations applied to the engagement subtypes were slightly modified to ensure that the items used to represent these constructs were applicable to students enrolled in college courses. The following sections detail the engagement aspects of the model that I will test, and the definitions that I applied to each engagement subtype.

2.9.2. Behavioral Engagement in the Current Study. Student involvement and participation are two components researchers typically use to assess student engagement across all levels of education (Astin, 1984; Fredericks et al., 2004; Kuh, 2009; Marks, 2000). Students' behavioral engagement with their course material in the current study is defined as the actions students place towards their academic studies. The definition I applied to this academic form of behavioral engagement differs from definitions applied by other researchers who use non-observable motivational indicators, such as persistence and effort during the assessment of student engagement (Marks, 2000; Newman, Wehlage, & Lambert, 1992; Skinner et al., 2009). While the current study relies on behaviors that are observable, it should be noted that this study relies on students' self-reported survey responses to assess all model components, including the engagement subtypes proposed in this study.

2.9.3. Emotional Engagement in the Current Study. The second engagement subtype in the proposed multidimensional engagement model is students' emotional engagement. Emotional engagement in the current study refers to students' attitudes or feelings towards various types of interactions that they encounter in their course. Aligning with Fredericks and colleagues' engagement framework, engagement is anticipated on being a multidimensional construct. Thus, all engagement components should be positively correlated with one another; however, I believe emotional engagement will share the strongest relationship with other engagement subtypes.

2.9.4. Cognitive Engagement in the Current Study. The final engagement subtype that I am proposing and testing in this study is students' cognitive engagement. I have defined cognitive engagement as the cognitive and metacognitive strategies that students apply to comprehend their course concepts and material. When examined in relation to students' course material, cognitive engagement is assessed through items that investigate the metacognitive strategies that students apply to learn their course material. Some researchers argue self-regulation is the highest form of cognitive engagement (Corno & Madinach, 1993; Meece, Blumenfeld, & Hoyle, 1988). Most assessments of self-regulation investigate the actions that students use to acquire information such as planning, monitoring, assessing knowledge, and regulating cognition (Patrick, Ryan, & Kaplan, 2007; Zimmerman, 1990).

For this dissertation, I applied indicators common during the measurement of self-regulation to assess cognitive engagement. I also wanted to determine if items that investigate students' perception of knowledge acquisition load consistently with items that investigated metacognitive learning strategies. Therefore, the items proposed to represent students' cognitive engagement investigate both students' metacognitive strategies and students' perceptions of knowledge acquisition. The engagement definitions that I recently detailed pertain to academic forms of engagement or students' engagement with their course material. The engagement model not only examines the three forms of engagement between students and their course content or course material, but items have also been developed that examine these forms of engagement between students and the different types of actors that students interact with during most, if not all, college courses

2.9.5. Student Interactions in the Current Study. A considerable amount of research has examined student interactions in online courses (Bernard et al., 2009). Bernard and colleagues (2009) conducted a meta-analysis on student interactions and provided the following definitions for the various forms of interactions that students typically encounter in online courses. *Student-content interactions* occur when students interact with their course material in order to comprehend the concepts presented. Students typically work independently on assignments during these types of interactions. *Student-to-student interactions* occur when students work in small groups or interact with other students on course related activities. Student-to-student interactions were largely absent when online courses were first developed and implemented, but advances in technology allow students to interact synchronously through web-based teleconferencing or asynchronously through mediums such as discussion boards or emails. These forms of interactions are believed to enhance students' comprehension of course material and concepts while also providing students with perceptions of community and peer support (Bernard et al., 2009; Rovai & Barnum, 2003). *Student-to-instructor interactions* occur when students interact with their instructors. These forms of interactions are believed to enhance students' understanding of course material, while also providing students with motivational support (Bernard et al., 2009). In the current study, items investigating student-instructor interactions have been modified to incorporate TAs, since teaching assistants may spend a significant amount of time with students in lieu of instructors throughout the duration of a course and provide students with motivational support.

2.9.6. Academic Forms of Engagement. Each engagement subtype in the proposed engagement measurement model will be examined in relation to the interactions students typically encounter in college courses. The first area of interaction in the proposed model pertains to students' engagement with their course content or material, terms which I use interchangeably throughout this dissertation. *Student-to-content behavioral engagement* refers to students' involvement, participation, and completion of classroom activities and assignments; *student-to-content emotional engagement* refers to students' perceptions of their course material and activities; and *student-to-content cognitive engagement* refers to students' comprehension of course material and the metacognitive strategies that students apply to learn the course material. Together these three areas reflect students' course engagement with their course material or simply academic forms of engagement.

2.9.7. Social Forms of Engagement. In addition to assessing academic forms of engagement, social forms of engagement are also included in the engagement model. Although the context in which engagement is examined differs across academic and social settings, I have defined each engagement subtype similarly. *Student-to-student behavioral engagement* pertains to the interactions that students experience with their classmates on course-related activities. Similar survey items were developed to measure the s*tudent-to-instructor behavioral engagement* construct; however, this model component refers to the interactions that students encounter with their instructor or teaching assistant. *Student-to-student emotional engagement* pertains to students' attitudes towards the interactions that they experience with their classmates on course-specific activities. Similarly, *student-to-instructor emotional engagement* reflects students' emotional reactions towards the interactions they experience with their instructors. The *student-to-student cognitive engagement* and *student-to-instructor cognitive engagement* constructs investigate the extent to which students' interactions with these course members increase their understanding of the course material and course content. Collaborating with instructors and classmates are expected to be highly correlated with the construct of students' course satisfaction, the final construct that will be tested via CFA in the proposed engagement measurement model.

2.9.8. Conceptualizations of Course Satisfaction in the Current Study. The final construct in the proposed engagement measurement is students' course satisfaction. Students' course satisfaction in this study pertains to students' perceptions of their entire course experience. While emotional aspects of engagement examine students' affective responses during specific activities, students' course satisfaction examines students' reactions towards their entire course experience. Since definitions applied to course satisfaction and emotional forms of engagement share similarities, I expect these model components to be highly correlated with one another. While it seems likely that a person who scores highly on measures of course satisfaction will also score highly on measures of academic and social engagement, it is possible that a student may be satisfied with the course but dislike specific course activities. Conversely, a person who rates a moderate level of course satisfaction may be report high levels of emotional engagement during specific learning activities or with specific course members. These nuances influenced my decision to include course satisfaction in the measurement model and test the validity of this construct through factor analytic techniques.

CHAPTER 3.0. METHODOLOGY

The overarching goal of this study is to establish a measurement model that is capable of comprehensively characterizing student engagement. In addition to establishing an engagement measurement model, I will test the final measurement model for invariance across the groups of students that I selected for this study. To achieve these goals, various types of factor analysis were performed. I began the analysis by first conducting a confirmatory factor analysis on the model recently proposed and described; however, the findings from this analysis did not meet accepted standards of fit. As such, I randomly split the entire dataset into two approximately equal halves, and I conducted an exploratory factor analysis (EFA) on the first half of the randomly split sample. I continued by performing a CFA on the same half of the dataset that I used to conduct the EFA to calibrate the model and determine the suitability of treating the model as second-order factor solution. After making two slight model adjustments, I proceeded to validate the model by conducting a CFA on the second half of the randomly split dataset. After establishing an appropriate engagement measurement model, I examined the second-order factor model for invariance by conducting several multiple-group CFA (MG CFA).

In this chapter, I provide some background on the sources of data obtained for this study, including the procedures that I used when collecting and handling the data and the characteristics of study participants. I continue this chapter by detailing the process of screening and assessing analytic assumptions as well as the findings from these analyses. I conclude this chapter by providing an overview of each type of analysis that I performed (i.e., EFA, CFA, and MG CFA) and detailing the analytic procedures for each of these analyses.

3.1. Study Sample & Participants

The study sample for the current study consisted of undergraduate students enrolled in various college courses that were offered at a premier public institution in the Western United States. Three of these courses were face-to-face courses, and the remaining 19 courses were online courses. Across all courses, 260 cases were removed from the entire study sample (N = 1,556), because participants did not agree to participate in the study; another 100 cases were removed, because students did not fully complete the post-course survey, which was the primary source of data used to assess model constructs. I retained 452 students from the face-to-face courses (37.8%) and 744 students from the online courses (62.2%), which together comprised the entire study sample (n = 1,196). These students were used to screen the data and assess analytic assumptions, which slightly lowered the total number of students used to conduct the final analyses for this dissertation.

3.2. College Courses

As previously noted, participants for the study were drawn from a total of 22 courses that were offered at seven different campuses. Most of the courses (81.8%) were offered at universities that operated under the quarter college system, while only four classes (18.2%) were offered at colleges that operated under the semester college system. The courses involved in the study were offered during five separate but consecutive academic terms, which first commenced during the spring 2012 academic term and culminated during the spring 2013 academic term. Six of the courses in the study were offered during the spring 2012 academic term, and two of these courses were offered in-person; four online courses were offered during the 2012 summer academic term, the fall 2012 academic term, and the spring 2013 academic term. During the winter 2013 academic term, four classes were also offered, but one of these four courses was a traditional, in-person course. Each face-to-face class had a comparable online course that was offered during the same academic term. Furthermore, fifteen of the courses (68.2%) were unique; whereas, the other seven courses (31.8%) were offered more than once during the academic terms from which data was collected.

3.3. Data Collection & Procedures

The data for the current study was initially collected to aid in completing an evaluation of an online course development program. There were various data sources obtained to conduct this evaluation; however, I only utilized two data sources for this dissertation. These forms of data consisted of survey and administrative or institutional data. In the following sections, I provide a more detailed description of these data sources

3.3.1. Survey Data. I initially planned to merge data from the pre-course survey, post-course survey, and administrative datasets and match these data sources by each case or student in the sample to create a single dataset for analysis. However, after I merged and matched these three data sources, there was a significant loss in student cases, particularly when merging pre-course survey data with post-course survey data. Fortunately, student demographic information, which I initially intended to obtain from the pre-course surveys, were provided on the administrative datasets. Therefore, analysis for the current study was conducted on a dataset that matched students' post-course survey responses to institutional data provided by university administrators. The post-course surveys were designed to gauge students' opinions and perceptions about online courses as well as understand students' course experiences. Themes that were examined in the post-course survey included pedagogical styles and students' approach to learning, comprehension of course content, course engagement, support seeking behaviors, and course satisfaction. All data used to test the model constructs are based on students' self-reported responses to post-course survey items that investigated these themes.

3.3.2. Institutional Data. Institutional data was obtained from university administrators and used to group students into different demographic categories. More specifically, I created subset datasets based on students' race/ethnicity and students' low-income status, and I used these datasets to conduct the tests of measurement invariance on the final engagement measurement model. The two variables that I used from this data source were students' racial or ethnic identity and students' low-income status. Students' race/ethnicity was the variable most commonly missing from these datasets. Of the 22 courses from which data was obtained for the study, students' ethnicity was missing from five of these courses, and students' low-income status was missing from three of these courses.

The administrative or institutional datasets grouped students based on Integrated Postsecondary Education Data System (IPEDS) reporting requirements. Therefore, I used definitions provided by IPEDS to categorize students into appropriate racial/ethnic groups. The datasets already contained a variable on students' low-income status, which was based on university thresholds. The definitions I consulted when categorizing students into different ethnic groups as well as the definition used by the university to identify low-income students are listed below in Table 1.

Table 1

Definitions used to Categorize Students into Ethnic/Racial and Low-Income Groups

Term	IPEDS Definition
Asian	A person having origins in any of the original peoples of the Far East, Southeast Asia, or the Indian Subcontinent, including, for example, Cambodia, China, India, Japan, Korea, Malaysia, Pakistan, the Philippine Islands, Thailand, and Vietnam.
Caucasian	A person having origins in any of the original peoples of Europe, the Middle East, or North Africa.
African-American or Black	A person having origins in any of the black racial groups of Africa.
(Hispanic)/Latino/a	A person of Cuban, Mexican, Puerto Rican, South or Central American, or other Spanish culture or origin, regardless of race.
Low-Income Students	Student who qualify for Pell Grants.

3.3.3. Data Merging & Appending Procedures. Data analyzed for this dissertation was located on multiple data files, so these files needed to be merged and matched by each case or student so that the post-course survey and institutional data appeared on a single file for each of the 22 courses. I performed all data handling procedures using IBM SPSS for Windows version 20.0. After removing cases for research non-consent and survey non-completion, merging and matching the data, and appending all datasets to obtain a single dataset, only three of these courses contained sample sizes large enough to conduct advanced quantitative analyses in their current form; therefore, I created subsample datasets based on the grouping variables of course format, student ethnicity, and students' low-income status. The sample sizes obtained through these data handling procedures allowed me to proceed with the covariance analyses with confidence.

3.3.4. Datasets. After appending the data from each course to obtain the entire study sample or Dataset 1 (*n* = 1,196), I created three subsample datasets based on the groups I identified for this study. In addition to each student in the study, Dataset 1 contained students enrolled in online courses (*n* = 744) and students enrolled in face-to-face courses (*n* = 452), since every case represented one of these course formats. I used Dataset 1 to create other subsample datasets based on students' ethnicity and students' low-income status, which I have titled Dataset 2 and Dataset 3, respectively. Dataset 2 (*n* = 732) consisted of students from Latino/a, African-American, Asian, and Caucasian ethnicities. Within this dataset, I created two composite variables to identify historically high-achieving students (i.e., Asian & Caucasian students) (*n* = 432) and students from ethnicities who have not traditionally had educational attainments rate comparable to Asian and Caucasian students (i.e., African-American and Latino/a students) (*n* = 300). Dataset 3 (*n* = 950) consisted of students who were identified by institutional data as being low income (*n* = 447) and students who were not classified as being low income (*n* = 503). I also randomly split Dataset 1 into two approximately equal halves to create Dataset 4 after the initial CFA did not support the proposed engagement model. I used the first half of Dataset 4 (*n* = 556) to conduct a follow-up EFA, and I used the second half of the dataset (*n* = 597) to validate the engagement model.

When conducting factor analysis and other structural equation modeling (SEM) analyses, larger sample sizes provide more stable results (MacCallum, Widaman, Zhang, & Hong, 1999). There a number articles on best practices in factor analysis, and the following guidelines or rules have been reported regarding sample sizes: a sample size of 100 is poor, 200 is fair, 300 is good, and over 500 is exceptionally good when conducting factor analysis; others have noted that there should be at least 5 cases per variable in the study, but a sample size of 100 may be adequate if the structure coefficients are high (Fabrigar et al., 1999; Henson & Roberts, 2006; Kahn, 2006; Reio & Schuck, 2014). Fortunately, the smallest dataset in this study contained 300 cases, and all other datasets contained at least 400 cases. As explained in the Results Chapter, these sizes were slightly lower after I screened the data and assessed the data for analytic assumptions, but the reduction in sample size was minimal. Prior to checking the data, however, I examined the characteristics of study participants for the entire sample and for each subsample.

3.4. Descriptive Statistics – All Courses

As seen in Table 2, the distribution of students' gender was nearly identical for the entire study sample; however, there were slightly more female students (51.7%) enrolled in these courses than male students (48.2%). Most students enrolled in these courses were either Asian or Latino/a. While Table 2 indicates the percentage of Latino/a students and Asian students were 22.3% and 19.8%, respectively, these percentages do not account for the number of missing cases (32.1%). Among respondents whose ethnicities were provided in the administrative datasets, Asian and Latino/a students accounted for approximately (42.1%) of all students. As previously mentioned several of the administrative datasets did not include students' ethnic identity, which explains why 32.1% of these responses were missing. Of all class standings, freshman students were least represented in the study sample (15.3%), followed by seniors (16.4%), juniors (18.5%), and sophomores

(20.7%), which indicates that students from all class standings enrolled in these courses. Most of the students (57.9%) enrolled in these courses could focus on their studies without the additional stress or burden of employment. More than one-third (34.1%) of these students worked between 0 and 20 hours per week, while a small portion of students (7.0%) worked 21 or more hours per week. When taking a closer look at the descriptive statistics across course formats, differences in the characteristics of students who enrolled in these courses emerged.

3.4.1. Descriptive Statistics of Students Enrolled in Online and Traditional Courses. Comparing demographic and background characteristics between students enrolled in online and face-to-face courses illuminated several differences between these groups of students. In the online courses, gender was quite evenly split, with slightly more male students (52.8%) enrolling in the online courses than female students (47.2%). These statistics were comparable to the representation of gender found across all courses. However, a higher percentage of female students (59.1%) enrolled in traditional, in-person courses than male students (40.5%). In terms of students' ethnicity, Asian students (26.1%) were the most represented ethnic group in online courses, while Latino/a students were the most represented ethnic group in the traditional, in-person courses (30.3%). Slightly more students from low-income backgrounds enrolled in the traditional courses (46.7%) than in the online courses (31.7%).

One interesting difference in enrollment characteristics pertained to the number of hours worked. As may be expected, a higher percentage of students who were employed decided to enroll in online courses. Among students enrolled in the traditional, in-person courses, slightly less than three-fourths of these students

(74.6%) were unemployed; whereas, more than half of the students (51.5%) enrolled in the online courses were employed. In addition, students enrolled in the online courses had a higher percentage of students who worked more than 20 hours per week. Approximately one out of every ten students (10.2%) who decided to enroll in the online courses worked 21 or more hours per week; conversely, a mere 1.7% of students enrolled in the traditional courses worked more than 20 hours per week. The flexibility of online courses and ability for students to complete course requirements at their own leisure is likely why online courses attracted such a larger share of students who were employed and worked full time. Table 2 also summarizes the descriptive statistics for these students.

3.4.2. Descriptive Statistics of Students Across Ethnic Achievement Groups. In addition to examining the descriptive statistics among students enrolled in online and in-person courses, I examined the descriptive statistics for all other groups of students selected for this dissertation. For sake of simplicity, I have labeled students from Asian and Caucasian ethnicities as "achievers" and students from African-American and Latino/a as "underacheivers" in Table 3. As seen in Table 3, when examining gender among all students from these four ethnicities, gender was about evenly split; however, there were slightly more female students (54.2%) than there were male students (45.5%). Similarly, among these students, students' college standing was approximately evenly distributed across freshman (11.7%), sophomores (17.8%), juniors (17.6%), and seniors (14.8%), but differences emerged when examining between-group differences.

The high-achieving ethnic group contained a much larger share of students from junior (24.5%) and senior class standings (20.4%) than did the group of students from ethnicities who have historically struggled academically. Among the high-achieving group, only 7.7 percent of students were juniors and 6.7 percent of students were seniors. Furthermore, the high-achieving ethnic group contained slightly more students who were classified as being low income (42.1%) than did the groups of students who were classified as being low income (37.0%). Quite more of these students also enrolled in online courses (60.9%) than in face-to-face courses (39.1%). When comparing descriptive statistics between these two groups of students, there were prominent differences.

The group consisting of African-American and Latino/a students contained a larger share of female students (64.3%) than the percentage of females in the group consisting of students from Asian and Caucasian ethnicities (47.2%). Among Asian and Caucasian students, the percentage of male students from these ethnicities (52.5%) slightly edged out the percentage of female students from these ethnicities (47.2%). Regarding course enrollment decisions, the Asian and Caucasian students were much more likely to enroll in online courses (73.4%) than were students from African-American and Latino/a ethnicities (43.0%). Employment rates were quite similar across these groups of students. Slightly more than one-quarter of students from the Asian and Caucasian ethnic groups (27.8%) were classified as being low income, whereas, slightly less than two-thirds (62.7%) of students from African-American and Latino/a ethnicities were classified as being low income.

3.4.3. Descriptive Statistics of Students from High- and Low-Income Backgrounds. The dataset that contained students who were classified as being low income and students who were not classified as being low income (n = 950) mirrored gender rates previously reported: there were slightly more female (53.9%) students than male students (46.0%). However, when I compared the dataset consisting of students from low-income backgrounds (n = 447) to the dataset consisting of students who were not low income (n = 503), as shown in Table 4, more female students (59.5%) came from low-income backgrounds than females who were not from low-income backgrounds (48.9%). Latino/a students accounted for the largest group of students from low-income backgrounds (35.3%), while only 10.7 percent of students not classified as being low income were identified as being Latino/a.

Asian students were the next largest ethnic group that were classified as being low income with 19.5% of these students. Asian students (22.1%) were also the largest ethnic group among students not classified as being low income. Surprisingly, low-income students had unemployment rates that were comparable to students not classified as being low income. Just less than 60 percent (59.5%) of students from low-income backgrounds were unemployed, which was nearly identical to the unemployment rates for students that were classified as not being low income (56.1%). Regarding course format, two-thirds of the students (66.4%) who were not low-income were enrolled in an online course. Conversely, slightly more than half of the students who were low-income were enrolled in an online course (52.8%) (see Table 4).

Table 2

Descriptive Statistics for the Entire Study Sample and for Students Enrolled in Online and In-Person Courses

	All Courses (*N* = 1,196)		Online Courses (*N* = 744)		In-Person Courses (*N* = 452)	
Characteristic	Frequency	Distribution	Frequency	Distribution	Frequency	Distribution
Gender						
Female	618	51.7%	351	47.2%	267	59.1%
Male	576	48.2%	393	52.8%	183	40.5%
Missing	2	0.2%	0	0.0%	2	0.4%
Total	1,196	100.1%	744	100.0%	452	100.0%
Ethnicity						
Asian	267	22.3%	194	26.1%	73	16.2%
Latino/a	237	19.8%	100	13.4%	137	30.3%
Caucasian or White	165	13.8%	123	16.5%	42	9.3%
Mixed	71	5.9%	22	3.0%	49	10.8%
African Am./Black	63	5.3%	29	3.9%	34	7.5%
Other	9	0.8%	9	1.2%	0	0.0%
Missing	384	32.1%	267	35.9%	117	25.9%
Total	1,196	100.0%	744	100.0%	452	100.0%
College Standing						
Freshman	183	15.3%	142	19.1%	41	9.1%
Sophomore	247	20.7%	198	26.6%	49	10.8%
Junior	221	18.5%	191	25.7%	30	6.6%
Senior	196	16.4%	183	24.6%	13	2.9%
Graduate	7	0.6%	7	0.9%	0	0.0%
Missing	342	28.6%	23	3.1%	319	70.6%
Total	1,196	100.0%	744	100.0%	452	100.0%
Income Status						
Not Low-Income	503	42.1%	335	45.0%	168	37.2%
Low-Income	447	37.4%	236	31.7%	211	46.7%
Missing	246	20.6%	173	23.3%	73	16.2%
Total	1,196	100.0%	744	100.0%	452	100.0%
Employment Hours						
Unemployed	692	57.9%	355	47.7%	337	74.6%
0-5 Hours/Week	95	7.9%	67	9.0%	28	6.2%
6-10 Hours/Week	119	9.9%	85	11.4%	34	7.5%
11-15 Hours/Week	98	8.2%	79	10.6%	19	4.2%
16-20 Hours/Week	97	8.1%	77	10.3%	20	4.4%
21-30 Hours/Week	40	3.3%	38	5.1%	2	0.4%
More than 30 Hours	44	3.7%	38	5.1%	6	1.3%
Missing	11	0.9%	5	0.7%	6	1.3%
Total	1,196	100.0%	744	100.0%	452	100.0%

Table 3

Descriptive Statistics of Students from Historically High- and Low-Achieving Ethnicities

	All Students (N = 732)		Achievers (Asian & Caucasian) (N = 432)		Underachievers (Latino & African-American) (N = 300)	
Characteristic	Frequency	Distribution	Frequency	Distribution	Frequency	Distribution
<u>Gender</u>						
Female	397	54.2%	204	47.2%	193	64.3%
Male	333	45.5%	227	52.5%	106	35.3%
Missing	2	0.3%	1	0.2%	1	0.3%
Total	732	100.0%	432	100.0%	300	100.0%
<u>College Standing</u>						
Freshman	86	11.7%	49	11.3%	37	12.3%
Sophomore	130	17.8%	81	18.8%	49	16.3%
Junior	129	17.6%	106	24.5%	23	7.7%
Senior	108	14.8%	88	20.4%	20	6.7%
Graduate	3	0.4%	1	0.2%	2	0.7%
Missing	276	37.7%	107	24.8%	169	56.3%
Total	732	100.0%	432	100.0%	300	100.0%
<u>Income Status</u>						
Low-Income	308	42.1%	120	27.8%	188	62.7%
Not Low-Income	271	37.0%	191	44.2%	80	26.7%
Missing	153	20.9%	121	28.0%	32	10.7%
Total	732	100.0%	432	100.0%	300	100.0%
<u>Employment Hours</u>						
Unemployed	426	58.2%	232	53.7%	194	64.7%
0-5 Hours/Week	54	7.4%	41	9.5%	13	4.3%
6-10 Hours/Week	78	10.7%	48	11.1%	30	10.0%
11-15 Hours/Week	54	7.4%	34	7.9%	20	6.7%
16-20 Hours/Week	49	6.7%	28	6.5%	21	7.0%
21-30 Hours/Week	29	4.0%	20	4.6%	9	3.0%
More than 30 Hours	34	4.6%	24	5.6%	10	3.3%
Missing	8	1.1%	5	1.2%	3	1.0%
Total	732	100.0%	432	100.0%	300	100.0%
<u>Course Format</u>						
Online	446	60.9%	317	73.4%	129	43.0%
In-Person	286	39.1%	115	26.6%	171	57.0%
Missing	0	0.0%	0	0.0%	0	0.0%
Total	732	100.0%	432	100.0%	300	100.0%

Table 4

Descriptive Statistics of Students Classified as Being Low Income and Students Not Classified as Being Low Income

	All Students (*N* = 950)		Not Low-Income (*N* = 503)		Low-Income (*N* = 447)	
Characteristic	Frequency	Distribution	Frequency	Distribution	Frequency	Distribution
Gender						
Female	512	53.9%	246	48.9%	266	59.5%
Male	437	46.0%	256	50.9%	181	40.5%
Missing	1	0.1%	1	0.2%	0	0.0%
Total	950	100.0%	503	100.0%	447	100.0%
Ethnicity						
Latino/a	212	22.3%	54	10.7%	158	35.3%
Asian	198	20.8%	111	22.1%	87	19.5%
Caucasian or White	113	11.9%	80	15.9%	33	7.4%
Mixed	57	6.0%	28	5.6%	29	6.5%
African American or Black	56	5.9%	26	5.2%	30	6.7%
Other	5	0.5%	4	0.8%	1	0.2%
Missing	309	32.5%	200	39.8%	109	24.4%
Total	950	100.0%	503	100.0%	447	100.0%
College Standing						
Freshman	166	17.5%	98	19.5%	68	15.2%
Sophomore	219	23.1%	121	24.1%	98	21.9%
Junior	153	16.1%	100	19.9%	53	11.9%
Senior	124	13.1%	74	14.7%	50	11.2%
Graduate	2	0.1%	2	0.4%	0	0.0%
Missing	286	30.1%	108	21.5%	178	39.8%
Total	950	100.0%	503	100.0%	447	100.0%
Employment Hours						
Unemployed	548	57.7%	282	56.1%	266	59.5%
0-5 Hours/Week	72	7.6%	44	8.7%	28	6.3%
6-10 Hours/Week	95	10.0%	46	9.1%	49	11.0%
11-15 Hours/Week	85	8.9%	46	9.1%	39	8.7%
16-20 Hours/Week	78	8.2%	40	8.0%	38	8.5%
21-30 Hours/Week	29	3.1%	20	4.0%	9	2.0%
More than 30 Hours	33	3.5%	19	3.8%	14	3.1%
Missing	10	1.1%	6	1.2%	4	0.9%
Total	950	100.0%	503	100.0%	447	100.0%
Course Format						
Online	571	60.1%	335	66.6%	236	52.8%
In-Person	379	29.9%	168	33.4%	211	47.2%
Missing	0	0.0%	0	0.0%	0	0.0%
Total	950	100.0%	503	100.0%	447	100.0%

3.5. Measures

I initially selected 47 continuous variables to represent one of the 12 latent constructs included in the proposed engagement model. After further reviewing these items, I removed two survey items or indicators, because they were only asked to students enrolled in online courses. The remaining 45 items were asked to every participant in the study, which was necessary since I am attempting to validate a model that is applicable to students enrolled in online and in-person college courses. In Table 5, I list each survey item underneath the construct that I forced them to represent during the initial CFA. I also list the variable code that was applied to each survey item or indicator, the initial scale for each survey item, and whether the item was reverse coded.

Nearly all items that I selected for the initial CFA were rated so that higher ratings reflected positive scores of the attribute being measured, which resulted in only having to reverse code one of the variables. Furthermore, the survey items were all rated on a 7-point Likert-Type scale. On all but three of the 45 indicators, the value of "1" meant students "Strongly Disagreed" with the statement, and the value of "7" indicated that students "Strongly Agreed" with statement or survey item. Among the three items not rated on this scale, the value of "1" represented "Never" and the value of "7" represented "Often". The following subsections summarize the variables that I tested to determine if they represented the latent constructs that I proposed in the engagement measurement model.

3.5.1. Pedagogical Approaches. I labeled the two pedagogical model constructs effective course design/instruction and collaborative learning. I selected four survey items to represent the effective course design/instruction construct and five different survey items to represent the collaborative learning, which examined instructors' promotion of collaborative learning in the course. All nine of these indicators were rated on a 7-point Likert-Type scale, ranging from "Strongly Disagree" to "Strongly Agree". The following statements reflect two of the four items that I believed represented the effective course design/instruction construct: "Class material was organized in way that made sense" (q_5_2); and "I knew what I needed to do for this course each week" (q_9_7). Items that investigated instructors' facilitation of collaborative learning asked students to indicate the extent to which they agreed that the course promoted a high level of interaction with various course actors. The last item asked to students to rate the extent to which they agreed with the following statement: "I was often given helpful feedback from the instructor or teaching assistant" (q_9_2).

3.5.2. Academic Forms of Engagement. I selected 14 survey items to represent distinct academic engagement subtypes. As previously noted, the academic engagement subtypes consisted of students' behavioral engagement with their course material, students' emotional engagement with their course material, and students' cognitive engagement with their course material. In Table 5, the abbreviation "SC" is written before each academic engagement subtype that is examined between students and their course content/material. I selected four items to represent the academic form of behavioral engagement; three items to represent the academic form of emotional engagement; and seven indicators to represent the academic form of students' cognitive engagement. These survey items were measured using the same 7-point Likert-Type scale that was used for the pedagogical model constructs.

3.5.3. Social Forms of Engagement: In addition to examining academic forms of engagement, I identified items that I believed loaded onto the six proposed social engagement constructs. The three engagement subtypes were examined in relation to students' interactions with their classmates as well as their interactions with their instructors/teaching assistants. These constructs are labeled students' behavioral, emotional, and cognitive engagement with their classmates and their behavioral, emotional, and cognitive engagement with their instructors/teaching assistants. In Table 5, I use the abbreviation "SS" before each engagement subtype that pertains to the student-student social form of engagement. I use the abbreviation "SI/TA" to represent the student-instructor/TA social form of engagement. These six social engagement constructs are represented by 19 indicators.

Students' behavioral and cognitive engagement with their classmates are each represented by three different survey items, and students' emotional engagement with their classmates is represented by four survey items. I selected four items for the student-instructor/TA behavioral engagement construct; three items for the student-instructor/TA emotional engagement construct; and two items for the student-instructor/TA cognitive engagement construct. The three items that contained the verbal anchors that ranged from "Never" to "Very Often" all related to social forms of behavioral engagement. These three items asked students to indicate the frequency in which they sought course support from various course actors, including their classmates (q_7_1), their teaching assistants (q_7_3), and their instructors (q_7_4). The student-student behavioral engagement construct also contained the only survey item that needed to be reverse coded so that positive ratings reflected positive scores on the item being measured. This item asked students to rate the extent to which they agreed with the following statement: "I felt isolated from my classmates (q_8_2)." I reversed coded the responses so that the value of "1" suggested that students agreed that they were isolated, and the value of "7" indicated that students strongly disagreed or did not feel they were isolated in the course.

3.5.4. Students' Course Satisfaction. The final latent construct that I tested in the engagement model was students' course satisfaction. I chose three survey items to represent this latent construct. One of these survey items directly measured students' course satisfaction by asking students to rate the extent to which they agreed with the following statement: "Overall, I was satisfied with this course" (q_19_3). The other two items tapped into areas that would indicate they were satisfied with the course. These additional items asked students to rate the extent to which they agreed with the following statements: "My attitude toward the subject matter improved as result of this course" (q_19_2), and "I would recommend this course to others" (q_19_4). I list each item that I selected and tested during the initial CFA underneath their respective construct in Table 5.

Table 5

Initial Variables and Associated Latent Constructs in the Proposed Engagement Measurement Model

Variable Per Latent Construct	Description	Initial Scale	Reverse Coded
Effective Course Design/Instruction			
q_2_2	This course was accessible anytime/anywhere.	1-7; SD - SA	NA
q_2_3	This course had a high-quality curriculum.	1-7; SD - SA	NA
q_5_2	Class material was organized in way that made sense	1-7; SD - SA	NA
q_9_7	I knew what I needed to do for this course each week.	1-7; SD - SA	NA
q_9_8	It was easy to find and access the work that I needed to do for this course each week.	1-7; SD - SA	NA
Collaborative Learning			
q_2_4	Course promoted a high level of interaction with classmates about course content.	1-7; SD - SA	NA
q_2_5	Course promoted a high level of interaction with teaching assistants about course content.	1-7; SD - SA	NA
q_2_6	Course promoted a high level of interaction with instructors about course content.	1-7; SD - SA	NA
q_9_2	I was often given helpful feedback from the instructor or teaching assistant.	1-7; SD - SA	NA
SC Behavioral Engagement			
q_9_1	There were many ways for me to check my understanding of the course material (e.g., quizzes, prompts, resources).	1-7; SD - SA	NA
q_9_5	I took advantage of the resources available in this course.	1-7; SD - SA	NA
q_9_9	I participated in all course assignments and activities.	1-7; SD - SA	NA
q_9_10	I completed all of my assignments by the due date.	1-7; SD - SA	NA
SC Emotional Engagement			
q_4_2	I am very interested in the subject area of this course.	1-7; SD - SA	NA
q_10_1	I enjoyed the course materials and/or activities.	1-7; SD - SA	NA
q_10_2	The course materials and/or activities sustained my interest.	1-7; SD - SA	NA

SC Cognitive Engagement			
q_4_4	I learned the basic concepts taught in this course.	1-7; SD - SA	NA
q_10_3	The course materials and/or activities made me rethink ideas that I had about course concepts.	1-7; SD - SA	NA
q_10_4	I found the course materials and/or activities to be intellectually challenging.	1-7; SD - SA	NA
q_10_5	The course materials and/or activities caused me to reflect on my understanding of the course content.	1-7; SD - SA	NA
q_10_6	I was able to connect the course content to information outside the course curriculum.	1-7; SD - SA	NA
q_10_7	The course material and/or activities helped me understand key course concepts and facts.	1-7; SD - SA	NA
q_10_8	The course material and/or activities helped me remember key course concepts and facts.	1-7; SD - SA	NA
SS Behavioral Engagement			
q_7_1	How often did you seek out support from students enrolled in this course for help with this course?	1-7; NE - VO	NA
q_8_2	I felt isolated from my classmates.	1-7; SD - SA	1-7; SA - SD
q_8_3	I often interacted with my classmates.	1-7; SD - SA	NA
SS Emotional Engagement			
q_8_1	I developed a connection with my classmates.	1-7; SD - SA	NA
q_8_4	I enjoyed my interactions with my classmates.	1-7; SD - SA	NA
q_11_1	My classmates valued my thoughts and opinions.	1-7; SD - SA	NA
q_11_6	I felt comfortable sharing my thoughts and opinions with my classmates.	1-7; SD - SA	NA
SS Cognitive Engagement			
q_11_3	I learned how to interact more effectively with classmates to enhance my learning.	1-7; SD - SA	NA
q_11_5	My classmates made me rethink ideas that I had about course concepts.	1-7; SD - SA	NA
q_11_7	My interactions with classmates increased my understanding of course material.	1-7; SD - SA	NA
SI/TA Behavioral Engagement			
q_7_3	How often did you seek out support from teaching assistants for help with this course?	1-7; NE - VO	NA
q_7_4	How often did you seek out support from instructors for help with this course?	1-7; NE - VO	NA
q_8_5	I often interacted with the teaching assistants.	1-7; SD - SA	NA
q_8_7	I often interacted with the instructor.	1-7; SD - SA	NA

SI/TA Emotional Engagement			
q_8_6	I enjoyed my interactions with the teaching assistants.	1-7; SD - SA	NA
q_8_8	I enjoyed my interactions with the instructor.	1-7; SD - SA	NA
q_11_2	The teaching assistants and/or the instructor valued my thoughts and opinions.	1-7; SD - SA	NA
SI/TA Cognitive Engagement			
q_11_4	I learned how to interact more effectively with the teaching assistants and/or the instructor.	1-7; SD - SA	NA
q_11_8	My interactions with teaching assistants and/or the instructor increased my understanding of course material.	1-7; SD - SA	NA
Course Satisfaction			
q_19_2	My attitude toward the subject matter improved as a result of this course.	1-7; SD - SA	NA
q_19_3	Overall, I was satisfied with this course.	1-7; SD - SA	NA
q_19_4	I would recommend this course to others.	1-7; SD - SA	NA

* *Abbreviation Definitions for Latent Constructs:*

SC Student-Content or Student-Material
SS Student-Student
SI/TA Students-Instructor/Teaching Assistant.

** *Initial Scales*

SD - SA Strongly Disagree - Strongly Agree
SA - SD Strongly Agree - Strongly Disagree
NE - VO Never - Very Often

3.6. Overview of Statistical Analyses & Analytic Procedures

After identifying items for the model, screening the data, and assessing the analytic assumptions, I proceeded to test the relationship between indicators and latent constructs by conducting a CFA on all survey items previously discussed. While all prior data handling procedures were performed using SPSS, I used Mplus Version 6.01 (Muthén & Muthén, 1998-2010) to conduct nearly all of the latent variable modeling for this study. Since the proposed model did not meet accepted standards of fit (see Results for a more detailed explanation), I continued my analysis by conducting an EFA in an attempt to establish an engagement model that was supported both theoretically and statistically; however, prior to conducting this EFA, I reviewed the post-course survey and identified seven additional survey items that I believed could possibly represent one of the 12 latent constructs. I included these indicators with the 45 indicators that I initially selected, which resulted in me conducting the EFA on 52 indicators. These additional variables along with the initial scale are listed below in Table 6. As detailed below, none of the items needed to be reverse coded.

Table 6

Additional Indicators Incorporated into the Follow-up Exploratory Factor Analysis

Variable Code	Description	Original Scale	Reverse Coded
q_2_1	This course was self-paced.	1-7; SD - SA	NA
q_2_7	This course promoted a high level of interaction with the course materials.	1-7; SD - SA	NA
q_4_1	It was important for me to learn the content in this course	1-7; SD - SA	NA
q_4_3	I understood the most difficult material presented in this course	1-7; SD - SA	NA
q_5_2	The format of this course allowed me the freedom to organize my time more effectively	1-7; SD - SA	NA
q_6_1	After this course, I plan to take more classes in this subject area	1-7; SD - SA	NA
q_6_2	I felt confident about my ability to perform well in this course.	1-7; SD - SA	NA

Note. *SD* = strongly disagree; *SA* = strongly agree.

After conducting the EFA, calibrating the model, and validating the final measurement model, I conducted tests of measurement invariance via multiple-group CFA (MG CFA) on each subsample to determine if these groups of students interpreted the model similarly. Testing the invariance of a factor model is a labor-intensive process that requires constraining various aspects of the model. I conducted seven different invariance tests across each of the groups that I selected for this study. Prior to detailing the results from these analyses, I provide an overview each analysis that I performed, while also detailing the statistical procedures that I employed during these analyses.

3.7. Factor Analysis

The primary goal of factor analysis is to determine the fewest number of latent constructs or factors that are able to account for the variance and covariance of a larger a set of measured variables or indicators (Brown, 2006; Henson & Roberts, 2006). Factor analysis allows researchers to examine the number, nature, and relation between factors that are used to represent the structure of correlations among a set of measured variables (Brown, 2006; Fabrigar, Wegener, MaCallum, & Strahan, 1999; Preacher & MacCallum, 2003). There are a range of techniques and approaches that may be applied when conducting these analyses, but factor analysis is classified into two broad categories—exploratory factor analysis (EFA) and confirmatory factor analysis (CFA).

One of the key similarities between EFA and CFA is that they are both based on the common factor model, which divides the variance of each indicator into common variance and unique variance (Brown, 2006; Preacher & MacCallum, 2003). Common variance refers to variance accounted for by common factors, and unique variance refers to variance that is not accounted for by common factors. Unique variance is further divided into variance that is specific to each variable and random-error variance. Random-error variance or simply random error occurs when some cause not associated with the latent factor is responsible for the variance, which may stem from issues with measurement, data collection procedures to problems, and/or the development of the variables (Brown, 2006; Floyd & Widamen, 1995; Preacher & MacCallum, 2003). The primary benefit of factor analysis is that it allows these sources of error to be identified. While these two approaches share similarities, there are key differences and specific reasons researchers would choose to conduct these forms of analyses

As the name implies, exploratory factor analysis is an exploratory procedure that is typically conducted when researchers do not have an *a priori* theory regarding the relationship between the variables and the latent constructs (Floyd & Widaman, 1995). During EFAs, variables are allowed to correlate freely with one another, which aids in identifying the most parsimonious model (Reio & Schunk, 2014). Exploratory factor analyses are typically performed during the early stages of research to eliminate variables that are not helpful in determining a final measurement model, develop or revise existing theories, and generate or adjust hypotheses about the theoretical processes (Reio & Schunk, 2014). An EFA may also be conducted when a hypothesized model is not supported through a CFA (Floyd & Widaman, 1995), which is the technique that I applied for this dissertation. While EFAs are typically performed when researchers do not have a strong theory regarding the relationship between variables and latent constructs, confirmatory factor analysis is used to explicitly test or confirm existing theory (Reio & Schunk, 2014). These theories are tested by imposing restrictions on variables that group specific variables together to represent latent constructs. The analytic procedures required to perform these analyses and evaluate the adequacy of models share some similarities, but they also slightly differ.

3.7.1. Evaluating Model Fit – Goodness-of-Fit Indices. During the evaluation of SEM models, there are a range of model or fit statistics that aid in identifying appropriate solutions. These model statistics are commonly referred to as *goodness-of-fit* indices. These indices are used to estimate the overall fit of a model and find values of parameters that replicate the covariance matrix as closely as possible (Brown & Cudeck, 1993; Floyd & Widaman, 1995). Researchers may apply several different factor extraction methods to obtain these indices, but the maximum likelihood (ML) estimation method yields the most comprehensive set of fit indices (Conway & Huffcut, 2003; Costello & Osborne, 2005; Fabrigar et al., 1999). Multivariate normality is an assumption associated with the ML method; fortunately, preliminary data screening suggested the data for the current study was normally distributed. Therefore, in the current study I used the ML method to estimate factors in all models that I tested, and I used fit indices associated with this estimation method, in part, to evaluate the adequacy of models.

Goodness-of-fit indices fall into several overarching categories, which each yield slightly different information regarding the extent to which a model is capable of replicating the initial the covariance matrix (Brown, 2006, Byrne, 2012). I used goodness-of-fit indices that are a part of the following categories during my evaluation of model fit: absolute fit, parsimony correction, and comparative fit. Absolute fit indices assess a model fit without taking into account other aspects that may influence the fit of the model, such as a model's fit in relation to a more restricted solution (Brown, 2006). The χ^2 and likelihood ratio are commonly used to assess model fit. This model statistics falls under the category of absolute fit. When the χ^2 test is nonsignificant, no other tests are typically conducted, and the model is accepted as being plausible or of decent fit (Kahn, 2006). However, this test is highly prone to error, especially when the sample size is large (Fabrigar, et al., 1999; Bentler & Bonnet, 1980). A small sample size may lead to nonsignificant findings, and a large sample size may lead to statistically significant findings (Marsh, Balla, and McDonald, 1988); therefore, I consulted several other indices when evaluating models.

Another index a part of the absolute fit category is the standardized root mean squared residuals (SRMR). The SRMR measures the difference between the observed variables and the hypothesized covariance matrix, while also adjusting for model complexity (Brown, 2006; Kahn, 2006). I also reviewed the root mean square residual (RMSEA), which falls under the parsimony correction category (Brown, 2006). Parsimony correction indices penalize models that have poor parsimony or more degrees of freedom than necessary. The last two indices that I used during my assessment fell under the comparative fit index, which evaluates the fit of a model fit in relation to a more restricted model. The comparative fit index (CFI) and Tucker-Lewis index (TLI) were the two comparative fit indices that I used to evaluate models in this study.

Attempts have been made to develop universal guidelines and cutoff values for assessing the adequacy of models (Hu and Bentler, 1999), but no one universal method has been concretely established (Marsh, Hau, & Wen, 2004). Instead, all fit indices should be examined together and coupled with a priori theory on how the variables are related to one another when making the final decision. Hu and Bentler (1999) encouraged researchers to use a 2-criterion or -index reporting strategy and argued the SRMR should be included at the very least. For this study, I consulted all of the indices recently detailed, which consisted of four indices from three different categories, and I used the cutoff values listed below as suggested by Hu and Bentler (1999).

1. SRMR values should be close to or less than .08 (i.e., SRMR $\leq$ .08);
2. RMSEA values should close to or less than .06 (i.e., RMSEA $\leq$. 06); and
3. CFI and TLI values should be close to or greater than .95 (i.e., $\geq$ .95).

It is important to note that these values are not absolute, and violations to one these fit indices do not necessarily mean a proposed model does not adequately fit the observed data. Instead, I examined all of these fit indices together, and I coupled these statistics with a priori theory on how the variables were anticipated on relating to one another when making my final decision on the suitability of a model. I also complemented these indices by reviewing and consulting several additional model statistics.

3.7.2. Evaluating Model Fit – Additional Model Diagnostics. In addition to examining the goodness-of-fit indices, I reviewed communalities, factor loadings, and modification indices to identify localized areas of strain or potential areas of model misfit (Brown, 2006). Communalities or multiple R^2 values refer to the amount of variance among indicators that are explained by common factors (Brown, 2006). Hooper, Coughlan, and Mullen (2009) noted that items that contain communality values $< .20$ should be removed from the analysis. Items with low communalities are likely to inaccurately measure the construct, contain low levels of reliability, and/or distort model findings (Fabrigar et al., 1999). I also examined the loadings for each item when conducting the EFA and CFAs for this study. I ensured all factor loadings $\geq .40$, since items that do not significantly load onto their hypothesized or proposed factors should be removed from the analysis (Netemeyer et al., 2003).

The last set of model statistics that I reviewed were modification indices (MIs). Modification indices are post-hoc or post-analysis adjustments that can be made to a factor model to improve its overall fit by reducing the model's χ^2 value (Brown, 2006). However, I solely used MIs to identify issues with model configuration, because implementing MIs are not recommended unless they are justified practically or theoretically (Kahn, 2006). I used these indices as well as the goodness-of-fit indices to evaluate all models that I analyzed in this study. However, there are number of additional critical analytic decisions and tests that must be performed when conducting EFAs and MG CFAs.

3.7.3. Exploratory Factor Analysis – Analytic Procedures. Exploratory factor analysis requires researchers to make critical decisions at different stages of analysis to obtain a reliable factor model, and these decisions rarely have absolute rules or guidelines (Costello & Osborne, 2005; Henson & Roberts, 2006). Researchers must select a factor estimation method to extract model constructs, determine the number of factors to retain in their model, apply a factor rotation method to aid in achieving simple structure, and interpret the latent construct(s) represented by the indicators (Floyd et al., 1986). The range of choices available to researchers add to the complexity of EFA and increase the likelihood of arriving at an improper solution. For instance, SPSS has eight different estimation or extraction methods (Costello & Osborne, 2005), while Mplus allows three different types of EFAs to be performed, which each have different factor estimation options based on the types of variables included in the analysis (Muthén & Muthén, 1998-2010). As previously noted, I used the ML method to extract factors in all models that I tested, because of the comprehensive set of indices yielded through this method. After determining the factor estimation method, researchers must determine the number of factors to retain or select. There are also multiple methods that may be used to make this determination.

I consulted Kaiser's Criterion, Catell's Scree Test, Parallel Analysis, and goodness-of-fit indices when selecting the number of factors to select in the model. Kaiser's criterion or the K1 rule serves as quick and simple technique for identifying factors to retain. During this method, the eigenvalues associated with each factor are reviewed, and factors that contain eigenvalues > 1.0 are retained. Conversely, factors with eigenvalues ≤ 1.0 are discarded from the model, because the variance that can be explained by one of the factors is less than the variance of a single indicator (Preacher & MacCallum, 2003). While this method provides an easy way of identifying the number of factors to retain, the Kaiser criterion is highly prone to overfactoring and occasionally prone to underfactoring (Fabrigar et al., 1999; Reio & Schuck, 2014). Therefore, I reviewed several additional factor selection methods.

I continued by conducting Catell's scree test, which tends to be more reliable than Kaiser's criterion (Hayton, Allen & Scarpello, 2004; MacCallum, Widaman, Zhang, & Hong, 1999). During Catell's scree test, each factor's eigenvalues are plotted in descending order against each factor, and the number of factors to retain is determined by identifying the location on the graph where the eigenvalues decline the greatest (Brown, 2006; Floyd & Widaman, 1995). While this approach is more reliable than Kaiser's criterion, Catell's scree test is subjective and may be difficult to interpret when there are multiple breaks or no clear discontinuities (Hayton et al., 2004; Preacher & MacCallum, 2003). Therefore, I also conducted a parallel analysis. Although this method is not widely used or reported, parallel analysis is one of the most accurate factor retention methods available (Reio & Schuck, 2014). Coupling this approach with Catell's scree test provides even greater reliability (Conway & Huffcutt, 2003; Ferguson & Cox, 1993; Ford, MacCallum, & Tait, 1986). There are two main types of parallel analysis, which are known as principal components parallel analysis (PCA-PA) and principle axis factoring parallel analysis (PAF-PA). Since I conducted a factor analysis for this study, I conducted a principle axis factoring parallel analysis.

During PAF-PA, eigenvalues associated with the data being analyzed are compared to eigenvalues that are associated with a random, simulated set of data

(Fabrigar et al., 1999; Ledesma & Valero-Mora, 2007). The eigenvalues associated with different factors for each dataset are then compared to one another to identify observed eigenvalues that exceed the randomly generated eigenvalues (Ledesma & Valero-Mora, 2007). To identify the number of factors to begin analyzing, I identified the number of factors associated with the observed eigenvalues that exceeded the simulated eigenvalues at the 95th percentile, which is more reliable than simply identifying observed eigenvalues and their factors that exceed the mean values of the simulated eigenvalues and their factors (Ledesma & Valero-Mora, 2007)[1]. I used these findings to identify the factor models, along with their associated fit indices and factor loadings, to begin analyzing, but I first compared several factor rotation methods to aid in obtaining interpretable solutions.

During EFA, factor loadings may be difficult to interpret if more than one factor is extracted (Reise, Waller, & Comrey, 2000), so researchers typically apply rotations to achieve "simple structure". Simple structure occurs when items load highly on only one factor and low on all other factors (Conway & Huffcutt, 2003; Ferguson & Cox, 1993). Factor rotations are classified into two broad categories, orthogonal rotations and oblique rotations. Orthogonal rotations do not allow

---

[1] The parallel analysis was the only aspect of the latent variable modeling analysis that I performed using IBM for Window SPSS version 20. I consulted the article and website below by O'Connor (2000) to obtain the syntax for the parallel analysis, while making several modifications to the syntax to reflect the data for this study: I changed the number of variables, selected the 95th percentile as the basis for comparing observed versus simulated eigenvalues; chose the principal axis/common factor analytic procedure; and generated permutations of the raw data, which is a more reliable method for identifying factors to retain. See the resources below for more information of conducting a parallel analysis.

O'Connor, B. (2000). SPSS & SAS program for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instruments, & Computers, 32*(3), 396-4020 https://people.ok.ubc.ca/brioconn/nfactors/rawpar.sps factors to correlate with one another; whereas, oblique rotations do allow factors to correlate (Floyd & Widaman, 1995). Since I expected a multidimensional engagement model with interrelated factors, I only compared oblique rotations. Furthermore, there appears to be little justification for implementing orthogonal rotations, because oblique rotations usually fit the data better than orthogonal rotations, allow factors to correlate, and also produce a solution that is equivalent to orthogonal rotations when factors are not correlated (Henson & Roberts, 2006; Preacher & MacCallum, 2003).

I applied and consulted the following rotations for the current study: geomin, direct oblique, promax, and quartimin rotations. The default rotation method for Mplus is geomin oblique rotation (Muthén & Muthén, 1998-2010), but this solution provided a fair share of cross-loadings. Furthermore, geomin rotations may not be the best rotation method to apply when anticipating complex factor models with three or more factors (Asparouhov & Muthén, 2009). Therefore, I continued to review additional rotation methods. The promax rotation did not provide the range of fit statistics associated with ML estimation method, so I opted not use this rotation for the analysis. The last two methods that I examined, direct oblimin and quartimin, yielded comparable model statistics and loadings. I based my final analysis on quartimin oblique rotations, but direct oblimin would have provided nearly identical findings. I used all previously detailed goodness-of-fit indices and model statistics to evaluate final model solutions. More specifically, during each round of analysis, I examined fit indices and identified problematic indicators that either contained significant cross-loadings, poor loadings on all factors, low communality values, high modification indices, and/or lack of theoretical justification. I then deleted these items from the analysis, calculated the new number of variables, conducted another parallel analysis, and repeated the analytic process by conducting another EFA. I repeated this process until I arrived at an engagement measurement model that was supported by theoretical and statistical standards, both at the individual item and at the overall model level. After establishing the model and validating the model on the second half of the randomly split dataset, I proceeded to test the model for invariance.

3.8. Measurement Invariance

Establishing invariance is a critical component of test or instrument development that may provide insight into biases of an instrument (Brown, 2006). An invariant measurement model allows researchers to determine whether between-group differences are due to actual character-trait or attitudinal differences of respondents or whether these differences are a result of variations in psychometric responses to survey items (Cheung & Renvold, 2002). Measurement invariance entails a series of tests to determine the extent to which aspects of a model function similarly across different populations, cultures, or time (Byrne, 2012; Schmitt & Kuljanin, 2008; Vandenberg & Lance, 2000). Instruments may be assessed for both measurement and structural equivalence (Byrne & Watkins, 2003), and parameters are tested in a logical order with increasingly restrictive constraints on model parameters to determine whether indicators and latent variables are interpreted similarly across groups. The parameters tested between groups during tests of measurement invariance include factor loadings, indictor intercepts, and residual variances (Brown, 2006; Byrne, Shavelson, & Muthén, 1989).

Within the CFA framework, two approaches for assessing invariance exist—multiple-group (MG) CFA and multiple indicators multiple causes (MIMIC) CFA. Multiple-group CFA and MIMIC CFA examine distinct facets of factor models to determine specific sources of invariance, but MG CFA is a much more comprehensive analytic approach. The primary advantage of conducting a MG CFA is that every source of measurement and structural invariance can be identified (Brown, 2006). During MG CFA, separate variance-covariance matrices for each group of interest are used to conduct simultaneous CFAs (Harrington, 2009). Measurement invariance is then assessed by conducting a series of CFAs in sequential or stepwise order and constraining different model parameters at each level or step of analysis (Brown, 2006; Byrne et al., 1989; van de Schoot, Lugtig, Hox, 2012). The constraints imposed on model parameters to test for measurement equivalency are known as equality constraints, which force different model parameters, such as the factor loadings, the indicator intercepts, and the residual variances, to contain equal values across groups (Brown, 2006; Byrne, 2012; Harrington, 2009). While MG CFAs allow sources of measurement and structural invariance to be identified, I solely focused on assessing the final engagement model for measurement invariance in the current study.

Muthén and Muthén (2009) noted that a model established through EFA and CFA procedures may be further studied to assess for invariance; thus, after I established the final measurement model through EFA and CFA procedures, I conducted a MG CFA to test for measurement invariance across the following groups of students:

1. Students enrolled in online courses and students enrolled in face-to-face courses;
2. Students from ethnicities who have historically performed well academically (i.e., Caucasian or White and Asian students) and students from ethnicities who have not performed as well as their racial counterparts (i.e., African-American or Black and Latino/a students); and
3. Students classified by university data as being low income and students not classified by university data as being low income.

In the following sections, I briefly explain all of the steps and tests that I performed to examine the final model for measurement invariance.

3.8.1. Measurement Invariance Procedures – Configural Invariance.

Invariance in a MG CFA framework first requires the measurement model to adequately fit in each group. Thus, the first step I conducted to examine measurement invariance across these groups was to determine if the model that emerged from the EFA and CFA adequately fit data for each group of students. If the measurement model does not fit, then the model is deemed to be non-invariant and invariance testing concludes. Not only must the model fit each group separately, but a CFA must be conducted on the two groups of interest simultaneously, which is known as tests of configural invariance.

Configural invariance, which is also known as the test of equal factor structure or the test of equal form, is considered the most basic level of measurement invariance (Brown, 2006; Chen et al., 2005; Vandenberg & Lance, 2000). Configural invariance is established if the number of factors and the item-factor relationship in the measurement model function similarly across groups

(Brown, 2006; Byrne, 2012; Chen, et al, 2005; Cheung & Rensvold, 2002; van de Schoot et al., 2012). Fit from this model must also adequately represent the data. The fit from this simultaneous analysis served as the baseline for next test of invariance. At each stage of invariance, I compared the fit from each newly constrained model to fit from the model conducted immediately before, which allowed me to determine whether the parameters constrained functioned similarly between groups.

3.8.2. Measurement Invariance Procedures – Metric Invariance. After testing the measurement model for configural invariance across all groups of interest, the next step consisted of testing for metric invariance, which is also known as the test of equality of factor loadings (Brown, 2006; Chen et al., 2005; Hirschfeld & von Brachel, 2014). The purpose of metric invariance is to examine the factor loadings across each group, which represents the strength of the linear relationship between each factor and their associated indicators (Chen et al., 2005). Metric invariance is established if the loadings of each indicator are equivalent across groups, which suggests that the members of each group construe the latent constructs or factors in the measurement model identically (Hirschfeld & von Brachel, 2014; Schmitt & Kuljanin, 2008; van de Schoot et al., 2012). Metric invariance allows the relationship between factors and indicators to be compared, because changes in scores on the latent factors are equivalent across groups (Dimitrov, 2010). I anticipated that a second-order measurement model would be needed to fit the data, which requires additional metric, scalar, and residual invariance tests. During the metric invariance tests of second-order factor models, all first-order and second-order factors are examined for invariance (Chen et al., 2005; Dimitrov, 2010).

There are two different metric invariance tests that must be conducted when second-order CFA models are being examined. During the first test of metric invariance, I constrained all first-order factor loadings to be equal across all groups, but I placed no other constraints on the model (van de Schoot et al., 2012). During the second metric invariance tests, in addition to constraining first-order factors, I constrained the factors a part of the second-order construct equal. Confirmation of invariance among first- and second-order factor models in second-order measurement models and the establishment or confirmation of scalar invariance are prerequisites for conducting cross-group comparisons. Groups that I deemed to contain equivalent first and second-order factor loadings were included in subsequent invariance tests.

3.8.3. Measurement Invariance Procedures – Tests of Scalar Invariance. Continuing with the tests of measurement invariance, I performed tests of scalar invariance. Scalar invariance, which is also known as strong factorial invariance, examines the equality of indicator intercepts or means across groups (Brown, 2006). As previously stated, at each level of invariance additional constraints are imposed. These constraints are incorporated simultaneously with all prior constraints made to the model. Thus, to test for scalar invariance, I further constrained the model that already contained first- and second-order factor constraints by fixing the means of each item or observed variable equal (Chen et al., 2005; Cheung & Rensvold, 2002; Hirschfeld & von Brachel, 2014). In addition to constraining the intercepts of observed variables, I set the factor intercepts/means of all first-order constructs equal. Scalar invariance is a prerequisite for comparing differences in factor means across groups (Chen et al., 2005). An invariant model at these levels suggests that differential item bias does not exist, because individuals who score the same on the construct obtain the same scores on indicators, regardless of their group affiliation (Milfont & Fischer, 2010). Furthermore, when scalar invariance is established across groups, scales of a measurement model contain the same origin and operational definition across groups and differences in latent means are result of true-group differences, which allows means on latent constructs to be compared across groups (Chen et al., 2005; Cheung & Rensvold, 2002; Dimitrov, 2009; Vandenberg & Lance, 2000).

3.8.4. Tests of Residual Variance Invariance. The final steps in assessing measurement invariance, in second-order factor models, are to examine the equality of factor disturbances and the equality of residual variances. Residual invariance tests are also known as strict factorial invariance. “Residual variance is the portion of item variance not attributable to the variance of the associated latent variable” (Cheung & Rensvold, 2002, p. 237). Thus, in addition to the previous model aspects that were constrained, I constrained the factor disturbances of all factors a part of the second-order factor to equality. I proceeded with this examination by testing the indicator or measured variable residuals, which I performed by fixing the residuals of each observed indicator equal. Establishing residual variance across groups suggests that the indicators used to measure latent constructs, and factors used to measure the second-order factor, contain the same degree of measurement error. This level of invariance suggests that the latent construct(s) are measured identically across groups (Cheung & Rensvold, 2002). Perhaps of most importance, establishing residual variance invariance suggests that differences between groups on the measured variables are a result of true latent factor differences (Widamen & Reise; 1997). Since invariance can be difficult to establish, researchers have argued for the examination of partial invariance when full invariance is not met (Byrne et al., 1989). At each step of analysis for each group of students, model modifications were examined across groups to determine if removing or adding parameters made substantive and theoretical sense.

3.8.5. Review of Measurement Invariance Tests: Altogether, I performed seven tests of invariance across all groups of students that I identified for this dissertation to investigate measurement invariance. Although investigating invariance is a lengthy process, especially when more than two groups are being compared, establishing measurement invariance is critical for subsequent analyses and for confidently making comparisons between groups. The list below summarizes the increasingly restrictive steps and tests that I performed to test for measurement invariance in the current study, beginning with the establishment of the engagement measurement model.

1. The CFA model that emerged from the EFA/CFA was specified;
2. This model was fit separately in each group via CFA;
3. Configural invariance was tested by fitting this CFA model simultaneously in the two groups being examined for invariance;
4. Metric invariance was assessed to determine equality of first-order and second-order factor loadings;
5. Scalar invariance was investigated to examine the equality of indicator intercepts;
6. Scalar invariance of first-order factor variances a part of the second-order factor model was conducted;

7. Factor disturbances of first-order factors a part of the second-factor model was assessed to determine equality first-order factor disturbances;
8. Residual variance invariance or strict factorial invariance was assessed to determine equality of residual variances among indicators;

3.8.6. Evaluating Invariance – Difference Tests. At each stage of invariance analysis, I compared fit from the newly constrained model to the fit associated with each preceding model in order to assess for invariance. Towards this end, I conducted the likelihood ratio test or the χ^2 difference test, which is commonly notated as $\Delta\chi^2$. The $\Delta\chi^2$ is one of the most frequently used approaches for testing competing SEM models[2] (Bryant & Satorra, 2012). Since all invariance models are nested, I used the findings from $\Delta\chi^2$, in part, to assess for invariance at each stage of analysis. During these tests, the χ^2 value and degrees of freedom or *df* of the less restrictive model are subtracted from the χ^2 value and *df* of the nested, more restricted model, which is the model being tested. The resulting χ^2 and *df* values are used to determine whether the parameter constraints significantly worsen model fit by using the χ^2 critical value table. A non-significant χ^2 value results in failing to reject the null hypothesis that the predicted covariance matrix is identical to the observed covariance matrix (Cheung & Rensvold, 2002). More simply, non-significant χ^2 values indicate that the imposed constraints do not significantly worsen

[2] To perform the $\Delta\chi^2$ the ML estimation method must be used and data must meet the assumption of normality. The $\Delta\chi^2$ cannot be used with estimators that adjust for missing or non-normally distributed data such as MLR and MLM; thus, Sattora & Bentler (1999) developed a scaled $\Delta\chi^2$ to address these issues. Muthén & Muthén detail the steps in their Mplus User Guide and on their website. For a practical guide on the Satora-Bentler scaled chi-square test see: https://www.statmodel.com/chidiff.shtml

fit and the model functions invariantly across groups (Bryant & Satorra, 2012; Cheung & Rensvold, 2002; Muthén & Muthén, 2009); however, a significant χ^2 value does not necessarily mean model aspects being tested are non-invariant (Cheung & Rensvold, 2002).

A major issue with using the $\Delta\chi^2$ and its associated χ^2 value to measure invariance is its sensitivity to sample size and model complexity (Cheung & Rensvold, 2002; Wu, Li, & Zumbo, 2007). Cheung and Rensvold (2002) tested 20 goodness-of-fit indices in a Monte Carlo simulation study and found that the only fit index not affected by the complexity of a model was the RMSEA. They suggested that the RMSEA should be used when assessing configural invariance. Their findings also confirmed that he $\Delta\chi^2$ may result in distorted findings with larger sample sizes. More importantly, their study revealed that ΔCFI is a more robust approach for assessing invariance across models (Cheung & Rensvold, 2002). A model is deemed invariant if the ΔCFI between the less restricted model and the nested model is less than .01. While ΔCFI values that are less than .01 indicate a model functions similarly across groups, ΔCFI among the more restricted model can improve in any value and still support invariance among the parameters being tested (Cheung & Rensvold, 2002; Dimitrov, 2010). Given these findings, I used both $\Delta\chi^2$ and ΔCFI at each stage of the analysis to assess for invariance.

3.8.7. Structural Invariance – Latent Mean Differences. Establishing measurement invariance or at least partial measurement invariance across groups of interest is a prerequisite for comparing latent means between groups (Byrne, 2012; Dimitrov, 2006); therefore, after conducting tests of invariance, I compared the latent mean scores on the final measurement model constructs across groups of students who interpreted the engagement model similarly. These analyses differed from my prior analyses in that they were based on mean and covariance structures (MACS); whereas, all previous analyses that I performed were based on the covariance structures (COVS) or matrices. In order to obtain latent mean scores of all first- and second-order model constructs, I requested two separate models for each of the groups of students.

I first obtained the broader academic engagement latent mean scores by constraining all factors in the model to equality in one of the groups, but allowing them to freely estimate in the other group. The group in which I constrained the constructs served as the reference group; while the exact latent mean scores cannot be determined, the strength and magnitude of the difference in latent mean scores between the groups being compared can be determined (Byrne, 2012). Therefore, all latent mean score estimates that I report indicate the difference in mean scores compared to the reference group (Brown, 2006). Furthermore, when examining latent mean scores of second-order factor models, latent mean scores of first-order constructs are not provided. Since I am interested in determining latent differences on any first-order engagement constructs, I will estimate a model as a first-order solution to identify latent mean differences between groups; therefore, I will compare both first- and second-order latent mean scores between groups who interpret the model invariantly.

CHAPTER 4.0. RESULTS

For this study, I conducted three main forms of analyses, which all expanded upon another and ultimately resulted in me fulfilling the overarching goals that I established for this study. The goals that I developed for the current study all necessitated specific forms of covariance or factor analysis. I first tested the model that I developed, which was primarily adapted from Fredericks', Blumenfeld, and Paris (2004) engagement framework, to determine the suitability of applying their multidimensional conceptualization of engagement to the assessment of student engagement in college course settings. I also incorporated literature on distance or online education to determine if the current data supported a model that examined engagement among the range of interactions that students encounter in college courses; thus, allowing engagement to be categorized into academic and social forms and specific sources of engagement to be identified. I also attempted to validate two constructs that related to pedagogy and a construct related to course satisfaction, which could possibly serve as antecedent and outcome variables in a theoretical model of engagement. Together, the model that I tested via confirmatory factor analysis contained 12 latent constructs

After determining that the proposed engagement model did not meet accepted standards of fit, I split the data into two approximately equal halves and performed an EFA of the first half of split dataset. Since the 12 latent constructs that I proposed were not supported by the data, I sought to establish an alternative model capable of characterizing student engagement. After identifying a factor solution potentially capable of measuring distinct forms of academic and social engagement, I calibrated the model. Towards this end, I conducted a CFA on the same first half of the dataset, and I pursued methods for improving the model prior to cross-validating it on the second-half of the dataset. After modifying the model, I proceeded to validate the model. I continued my analysis by examining the model for invariance across the following groups of students:

1. Students enrolled in online courses and students enrolled in face-to-face courses;
2. Students from ethnicities who have historically performed well academically (i.e., Caucasian or White and Asian students) and students from ethnicities who have not performed as well as their racial counterparts (i.e., African-American or Black and Latino/a students); and
3. Students classified by university data as being low income and students not classified by university data as being low income.

After establishing invariance across groups in which the model functioned invariantly, I completed my analysis by comparing the latent mean scores on all final engagement model constructs across the groups of students in the study. This results sections follows the order in which I conducted each analysis. I begin by summarizing the preliminary data screening that I performed and the checks of analytic assumptions. I proceed by detailing the findings from the initial CFA on the proposed engagement model. I then detail the findings from the EFA. I conclude this chapter by explain the results from tests of measurement invariance and comparison of latent mean scores.

4.1. Preliminary Data Screening

After preparing the datasets for analysis, creating grouping variables and subsample datasets, and perusing and selecting survey items to represent the 12 latent constructs based on theoretical and empirical literature, I used IBM SPSS for Windows version 20.0 to screen the data and assess assumptions associated with the analyses that I performed. The preliminary data screening assessments consisted of reviewing the data files for accuracy, examining missing responses, and determining the variation in responses to help identify respondent bias. I also checked the data to ensure that the assumptions associated with sample size and missing data, normality, linearity, outliers, and multicollinearity and singularity did not jeopardize the findings[3]. Meeting assumptions of factor analysis, which is a form of structural equation modeling, is critical to these types of statistical analyses, because violations to these assumptions are likely to result in inaccurate test statistics and/or Type I or Type II errors (Brown, 2006; Nimon, 2012).

I checked the quality and accuracy of the data by examining frequencies, standard deviations, distributions, and range of values for all continuous variables. Twenty cases contained the value of "0" on at least one indicator, which was not a valid response option. I replaced these values with missing case scores. A select number of students provided identical responses to each survey item. To determine cases with little to no variance in their responses, I transferred each case into an Excel spreadsheet and calculated the standard deviation for all survey responses. Six cases had a variance of 0.0 for the entire survey, and 12 additional cases contained variances $\geq .50$. I elected to remove these 18 items from the study sample. To further ensure the accuracy of the data, I randomly compared the values from the

---

[3] For a more thorough review of the analytic assumptions associated with factor analysis as well as solutions to commonly encountered issues with the data consult Tabachnick and Fidell (2007).

raw data files to values uploaded into SPSS. All items were correctly uploaded into the analytic program.

I continued my assessment by examining the frequency and patterns of missing cases. None of the items that I selected to represent model constructs contained missing values that exceeded 5 percent. Using the Missing Value Analysis function available in SPSS, I implemented an expectation maximization technique to conduct Little's MCAR $\chi 2$ Test, which was nonsignificant: ($\chi 2$ = 3765.29; *df* = 3714; *p* = .274); thus, I failed to reject the null hypothesis that the data was missing completely at random. After checking the integrity of the data, I proceeded to assess the analytic assumptions associated with SEM.

4.1.1. Analytic Assumptions. Proceeding with assessing the analytic assumptions, I first sought to determine whether the continuous variables selected for analyses were normally distributed. Univariate normality, as may be expected, is a condition that must be met before assuming variables adhere to multivariate normality (Burdenski, 2000). To assess for univariate normality, I examined the probability plots of each variable, which suggested that the data was indeed normally distributed. I then reviewed the skewness and kurtosis statistics for each variable. A variable is deemed to be normally distributed if the skewness and kurtosis have values around zero (Tabachnick & Fidell, 2007). Skewness variables that do not exceed $|2.0|$ and kurtosis values that do not exceed $|7.0|$ are deemed to be normally distributed (Fabrigar et al., 1999). None of the skewness or kurtosis values for any of the indicators deviated substantially from normality, which supported the visual findings from the probability plots. Establishing multivariate normality is quite burdensome, so researchers have contended that upon confirming univariate normality, multivariate normality is reasonable to assume (Floyd & Widamen, 1995; Warner, 2008). These findings indicated that variables that I selected for analysis met univariate normality, and the assumption of multivariate normality was also reasonable to assume.

Another assumption of SEM is that the data should not contain univariate nor multivariate outliers. An outlier occurs when a case has an extreme value on one or more variables (Tabachnick & Fidell, 2007). I tested variables for potential outliers on all indicators by first requesting and examining box plots for each survey item. My visual examination of these box plots revealed that there was a total of 16 outliers on four different variables that represented 13 different cases. Ten of these cases contained extreme values on only one variable; however, three of these variables were multivariate outliers, because they each contained extreme values on at least two different variables. After examining each case that contained outliers, I determined that these outliers existed because the responses provided by students were outside of normal distribution for these variables. Instead of simply deleting cases with these outliers, I carefully reviewed students' survey responses. I deleted six of these 13 cases, because participants provided inconsistent responses across similarly worded survey items

After checking the dataset for univariate outliers, I used Mahalanobis Distance to examine the entire dataset for multivariate outliers, which refers to the distance of a case from the intersection of the means for all variables (Tabachnick & Fidell, 2007). Using the number of variables selected for analysis as the degrees of freedom, I used the $\chi 2$ critical value table to identify the cutoff value associated with 45 degrees of freedom and a critical value probability less than the $\alpha = .001$, which was 61.66. Based on this value, 106 cases were identified as containing multivariate outliers, bringing the total number of cases with outliers in the dataset to 125. Instead of simply deleting these cases, I tested the final engagement measurement model on datasets with and without the outliers to determine if the outliers effected the final solution.

Continuing with the assessment of assumptions, I examined the linearity among the independent variables. To examine linearity, I requested and reviewed partial regression plots for each independent variable. The absence of curvilinear relationships and the approximately linear illustration that emerged through these partial regression plots indicated the assumption of linearity was achieved. The last assumptions I checked were those of multicollinearity and singularity, which cause issues with data when variables are very highly correlated with one another (Tabachnick & Fidell, 2007). Multicollinearity is present when variables are highly correlated, and singularity arises when variables are redundant (Tabachnick & Fidell, 2007). To determine whether multicollinearity or singularity existed in the current dataset, I examined the correlations between all variables as well as collinearity diagnostics.

I first requested and perused correlations for all variables selected for analysis. One set of variables contained correlations that exceeded .90. These two items asked students to rate their level of agreement with the following statements (1) The course materials and/or activities sustained my interest; and (2) I enjoyed the course materials and/or activities, which I believed would represent students' emotional engagement with their course material. While these two items were highly correlated, I continued to examine collinearity diagnostics associated with each variable. Researchers typically use the VIF value of 10 as a sign of severe multicollinearity (Hair, Ringle, & Sarstedt, 2011; O'Brien, 2007). None of the VIF values among the current model posed any issues, since none of these values contained VIF values above 10. The items that were expected to have VIF values above 10 were the two highly correlated items; however, the highest VIF value among these variables was 7.43. While the value of 7.43 suggests there may be slight multicollinearity issues with several of the variables selected for the analysis, these statistics did not exceed the VIF value of 10; thus, I concluded that data analyses could proceed with confidence.

4.2. Revised Sample Sizes

After conducting the preliminary data screening and assessing the assumptions for analyses, the size of each sample and subsample slightly differed from the sizes previously reported. The cases removed after conducting the preliminary data screening resulted in slightly lower sample sizes for each dataset and subsample dataset. Although removing cases after screening the data and assessing the analytic assumptions resulted in slightly smaller sample sizes, most of the data was preserved: only 43 total cases were removed from Dataset 1, 31 cases were removed from Dataset 2, and 36 cases were removed from Dataset 3. Initial and final sample sizes for each dataset are provided below in Table 7 along with the descriptive statistics. It is important to note that the observable data is all comparable before and after removal as would be expected given the modest reduction in sample size. I also provide the sample sizes for the dataset that I randomly split in half to perform the follow-up EFA and model validation after determining that the initial CFA did not meet accepted standards of fit (see Table 7).

Table 7

Summary of Initial and Final Dataset Sample Sizes

Characteristic	Initial Sample Size		Revised Sample Size	
	Frequency	Distribution	Frequency	Distribution
Dataset 1 - Course Format Cases				
Online Cases	744	62.2%	721	62.5%
In-Person Cases	452	37.8%	432	37.5%
Total (Entire Study Sample)	1,196	100.0%	1,153	100.0%
Dataset 2 - Ethnicity Cases				
Asian & Caucasian Students	432	59.0%	413	58.9%
Latino/a & African-American Students	300	41.0%	288	41.1%
Total (Entire Subsample)	732	100.0%	701	100.0%
Dataset 3 - Economic Status Cases				
Not Low-Income Students	503	52.9%	489	53.5%
Low-Income Students	447	47.1%	425	46.5%
Total (Entire Subsample)	950	100.0%	914	100.0%
Dataset 4 - Randomly Split Cases				
First Half of Data	-	-	570	49.4%
Second Half of Data	-	-	583	50.6%
Total (Entire Study Sample)	-	-	1,153	100.0%

4.3. Initial Confirmatory Factor Analysis

The multidimensional engagement measurement model that I proposed and tested during this analysis did not meet accepted standards of fit: $\chi^2(928) = 12006.19$, $p < .001$; CFI = .76; TLI = .74; RMSEA = .10 [.10; .10]; and SRMR = .11. While I applied different model estimators and analyzed the findings associated with multiple estimation methods, none of them significantly improved model fit. As noted in the methodology chapter, I used the ML estimator during this assessment as well as all model assessments throughout the study to extract factors. I also removed all outliers identified, and I tested the model again to determine if these significantly impacted the findings: $\chi^2(928) = 10905.78$, $p < .001$; CFI = .76; TLI = .75; RMSEA =

.10 [.10; .10]; and SRMR = .11 All fit indices that I reviewed were well above accepted guidelines provided by Hu and Bentler (1999) for datasets with and without outliers. There was a slight reduction in the χ^2 estimate, but all other fit indices were nearly identical. Since the CFAs that I conducted did not support the existence of this complex factor model, I continued my analysis randomly splitting the entire dataset into two approximately equal halves and conducting a follow-up EFA on the first half of the dataset.

4.4. Exploratory Factor Analysis

Prior to conducting the EFA in Mplus, I performed a parallel analysis on the same dataset using SPSS. Based on the PAF-PA, seven eigenvalues were above 1.0, and all of the principal axis eigenvalues were greater than the eigenvalues of the randomly generated eigenvalues at the 95th percentile. While the principle axis eigenvalues suggested that seven factors should be selected, the plotted versus random eigenvalues were not as clear cut (see Figure 3). The random versus plotted eigenvalues intersected at what appears to be 10 or 11 factors, which may be because the plot includes predicted and observed eigenvalues that are less than 1.0. After conducting the PAF-PA, I analyzed other factor selection methods. The scree plot associated with the data (see Figure 4) was also examined and used to conduct Catell's Scree Test, which suggested retaining eight factors. In addition to these analyses, I examined Kaiser-Gutmann's Criteria or the K-1 rule, which also supported an 8-factor model, therefore, I began analyzing an eight-factor model to identify items that contained high, low, and/or multiple cross loadings.

After identifying the number of factors to select and associated model statistics and fit indices to begin analyzing, I examined the quartimin rotated loadings associated with an 8-factor model solution to identify indicators that contained high, low, or multiple cross loadings. I also considered the literature that I reviewed when making final decisions to retain or discard items. During the first round of this EFA, I removed a total of 11 items from the analysis. All of these items contained moderate cross-loadings onto at least two factors, and most of the items loaded moderately onto more than two factors. After removing these 11 items, I conducted another PAF-PA and EFA, and I repeated the analytic process of identifying items to include and remove from the model.

4.4.1. Second Round EFA Factor Selection & Analysis – Findings. The PAF-PA indicated that six of the observed eigenvalues exceeded the randomly generated eigenvalues at the 95th percentile. Similar to the first analysis, the plotted eigenvalues suggested slightly more factors should be extracted (see Figure 5). The observed eigenvalues intersected with the random eigenvalues at approximately 10 eigenvalues; however, this plot may be misleading, since after the sixth factor, all additional data eigenvalues were less than 1.0. Zwick (1986) reported that components that have eigenvalues less than 1.0 must have at least three indicators to be of any interest to researchers, and even with three indicators, these components may not be useful. Catell's scree test (see Figure 6) and the K1 rule both suggested that that 8 factors should be selected; however, I decided to begin examining a ten-factor model solution, because the model statistics for the ten-factor solution provided the best fit: $\chi^2(455) = 1454.68$, $p < .001$; CFI = .949; TFI = .908; RMSEA = .063 [.06; .067]; and SRMR = .019. The fit statistics for the other solutions were well below standards needed to accept a solution with confidence.

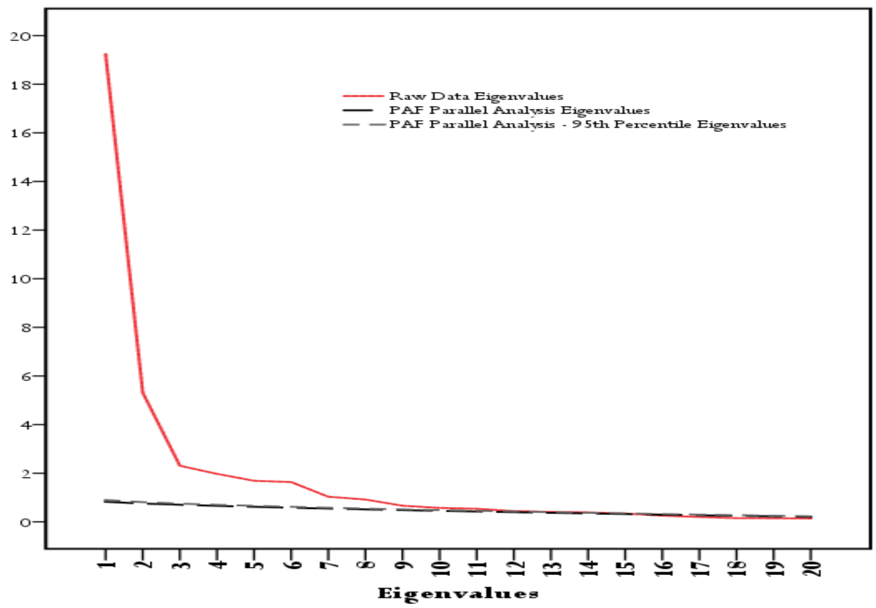

Figure 3. Plotted principle axis eigenvalues of observed versus randomly generated principle axis eigenvalues that transpired during the first round of principle axis factoring parallel analysis, including randomly generated eigenvalues at the mean and 95th percentile.

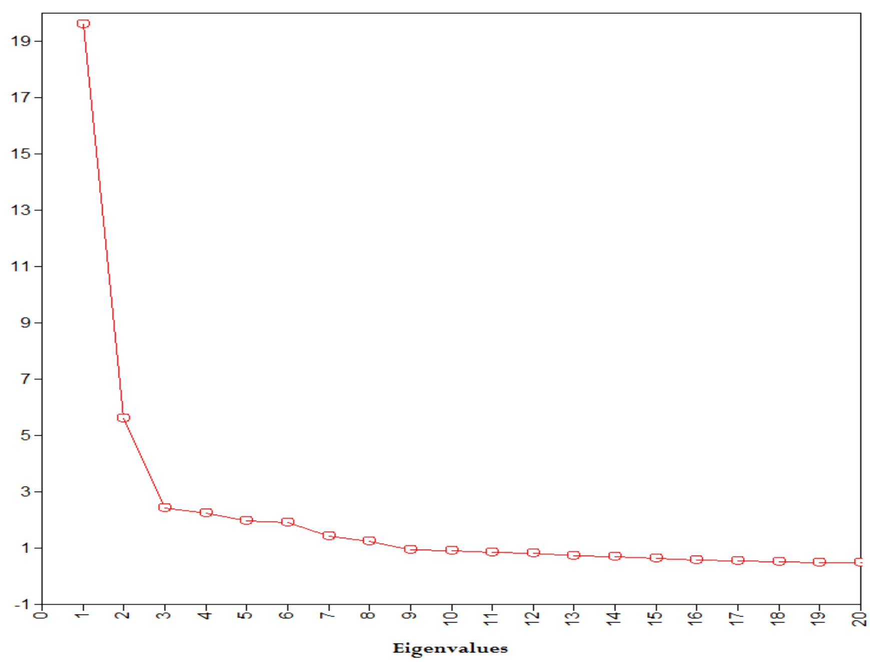

Figure 4. Scree plot of observed eigenvalues used to conduct Catell's scree test for the first EFA round.

I continued my EFA assessment by reviewing the quartimin rotated loadings of a ten-factor solution. After reviewing the rotated loadings, a more pronounced factor model started to become apparent. However, I still identified and removed six indicators during this round of analysis. Many items that I discarded during this round of analysis did not load strongly onto any single factor, while others contained moderate loadings on at least two other factors. This round of analysis revealed that the pedagogical constructs that I proposed were unlikely to be represented by the current data. After identifying these items for removal, I conducted a third round of analysis on the remaining 35 variables.

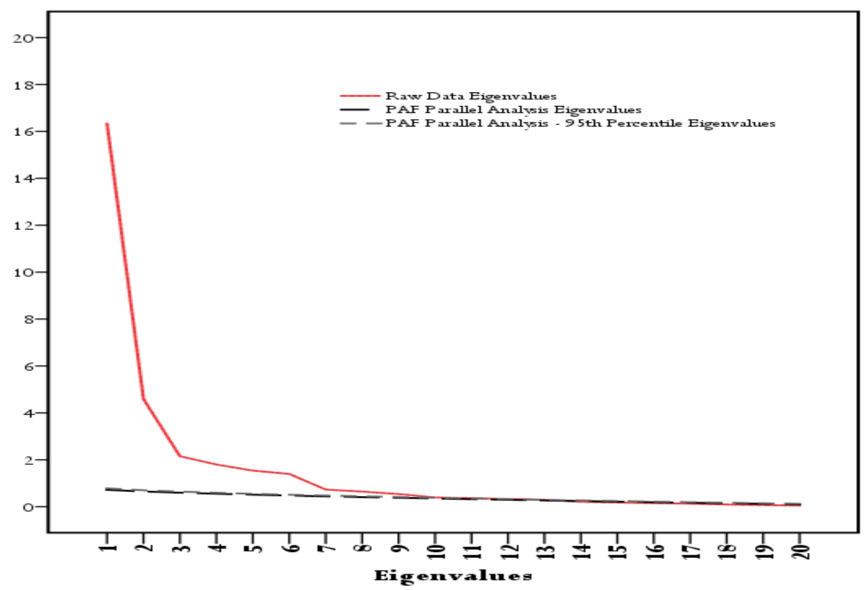

Figure 5. Plotted eigenvalues of observed versus randomly generated eigenvalues used during the second-round parallel analysis, including randomly generated eigenvalues at the mean and 95th percentile.

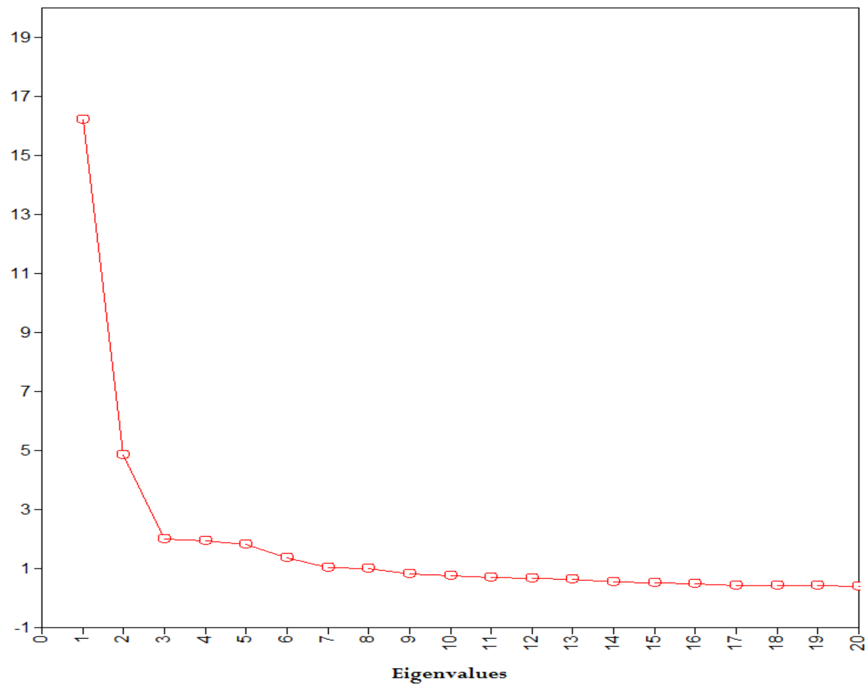

Figure 6. Scree plot of observed eigenvalues used to conduct Catell's scree test for the second EFA round.

4.4.2. Third Round EFA Factor Selection & Analysis – Findings. The PAF-PA that I performed for third round of analysis indicated that five factors should be retained (see Figure 7). Catell's scree test, illustrated in Figure 8, and Kaiser's criterion both suggested retaining a six-factor model solution. Since these factor selection analyses supported a six- or seven-factor model, I examined the fit statistics associated with a six-, seven-, and eight-factor model. The only fit statistics that approached adequate fit were for the eight-factor model: $\chi^2(343) = 1337.17$, $p < .001$; CFI = .942; TFI = .900; RMSEA = .073 [.068; .077]; and SRMR = .023. There was only a slight improvement between the fit statistics for this factor model and the factor model with nine factors, so I decided to begin analyzing statistics associated with an eight-factor solution.

During this third round, I removed a total of 17 items from the analysis. The removal of these items indicated that several key constructs initially proposed during the CFA were not well represented by the data. Items that I initially believed would represent effective course design, collaborative learning, student-instructor/TA engagement, and course satisfaction were all removed during the series of EFAs that I performed. The remaining items all loaded strongly onto their associated factor, and all items contained loadings that were greater than .40. After I performed these EFA iterations, I was left with a five-factor model. These factors included the three engagement subtypes between students and their course material (i.e., behavioral, emotional, and cognitive). In addition, two forms of student-student engagement transpired. To ensure that these items should be retained, I performed one last EFA.

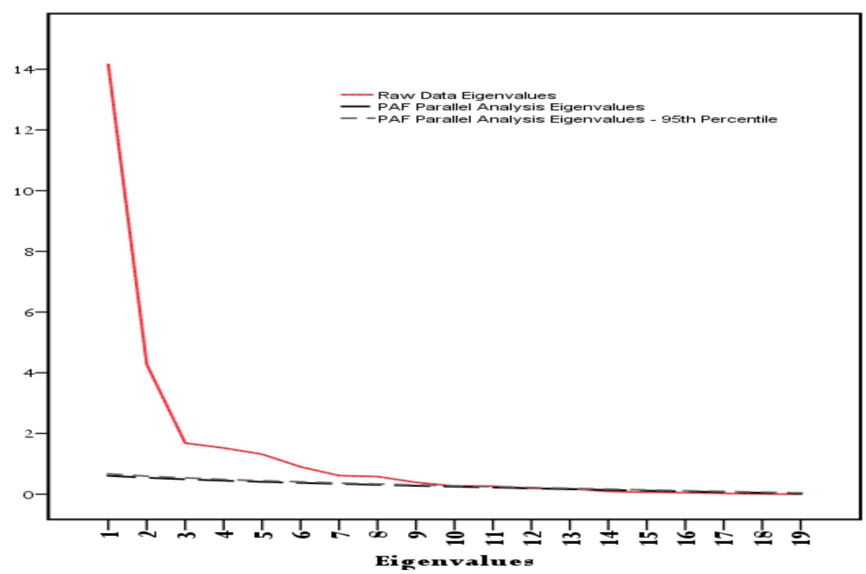

Figure 7. Plotted eigenvalues of observed versus randomly generated eigenvalues used during the third round of parallel analysis, including randomly generated eigenvalues at the mean and 95th percentile.

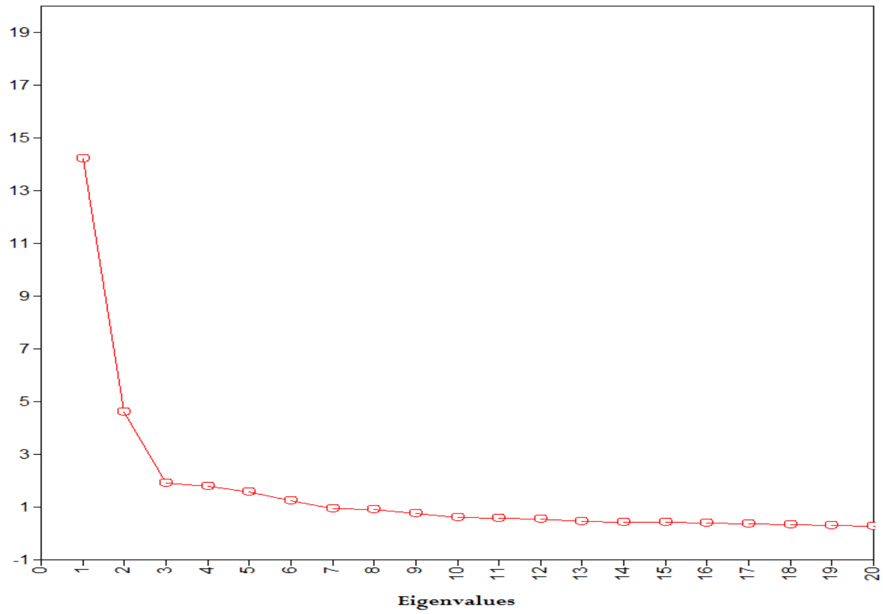

Figure 8. Scree plot of observed eigenvalues used to conduct Catell's scree test for the third EFA round.

4.4.3. Final EFA & Model Summary. The EFA resulted in a five-factor engagement measurement model that contained both academic and social engagement constructs. During all three EFA rounds, I deleted 34 of the 52 initial indicators that were included in the first EFA round, which resulted in 18 items representing five different factors. To confirm these findings, I performed one last EFA. Since findings from the last analysis indicated that the remaining 18 variables represented five different constructs, I started this analysis by examining the fit and model statistics associated with a five-factor solution. The fit statistics associated with this model confirmed that this was a plausible model: $\chi^2(73) = 388.70$, $p < .001$; CFI = .96; TLI = .92; RMSEA = .089 [.080; .097]; and SRMR = .024. While the RMSEA value was slightly above the recommended value, all other fit statistics strongly supported the model. Furthermore, all factor loadings exceeded .40, and most items contained moderate loadings above .60.

The EFA provided initial support for the measuring engagement through the three academic engagement subtypes proposed in this study (i.e., behavioral, emotional, and cognitive engagement). Students' behavioral engagement with their course material was represented by four different indicators, and the academic forms of emotional and cognitive engagement were each represented by three different indicators. In addition to these academic engagement subtypes, a social form of emotional engagement and cognitive engagement emerged, which both pertained to students' interactions with their classmates. Student-student cognitive engagement was represented by five items, and student-student emotional engagement contained three items (see Table 8). The relationship between variables and factors in this final model shared many similarities with the model that I initially proposed and tested during CFA, but several items operated differently than I expected.

I initially believed the first two items listed in Table 8, "q_9_8" and "q_9_7", measured the pedagogical construct of effective course design, but these items loaded well onto the academic form of behavioral engagement. The item labeled "q_9_8" asked students to rate the extent to which they agreed that they could find and access coursework each week, and the item labeled "q_9_7" asked students whether they were aware of what they needed to do for the course each week. Since students would need to be involved in their course to provide favorable responses to these items, and the EFA suggested these items loaded well with other items that measured students' behavioral engagement with their course material, I retained these items and forced them to represent this construct during subsequent model calibration and validation analyses. Every item that I included in the initial CFA to measure students' emotional and cognitive engagement with their course material emerged to represent these constructs during the final EFA model. The only difference between my initial theory and the EFA findings regarding the academic forms of engagement was that the academic form of cognitive engagement contained four fewer indicators.

The student-student cognitive engagement construct contained two items that I initially used to represent students' emotional engagement with their classmates. These items asked students to rate the extent to which they agreed with the following statements: "I felt comfortable sharing my thoughts and opinions with my classmates" (q_11_6), and "My classmates valued my thoughts and opinions" (q_11_1). Research has indicated that affect and cognition share many similarities and these measures often overlap (Dai & Sternberg, 2004), which increased my confidence in retaining these items and labeling this construct student-student cognitive engagement. Other items associated with this construct investigated the extent to which students' interactions with their classmates increased students' understanding of the course material (q_11_7) and whether these interactions caused them to reflect on course concepts (q_11_5), which were both measures that I initially proposed and tested.

I named the final factor in the model student-student emotional engagement. The following two items in this construct coincided with the model that I initially proposed and tested: "I developed a connection with my classmates" (q_8_1), and "I enjoyed my interactions with my classmates (q_8_4). The last item in this construct contained an item that I originally thought reflected students' behavioral engagement with their classmates. This item asked students to rate the extent to which they agreed with the following statement: "I often interacted with my classmates" (q_8_3). It is likely that students who often interact with their classmates will develop a bond or connection with these members, which is why I deemed it was appropriate to test this item as an indicator of the student-student emotional engagement construct.

4.5. Model Calibration

After identifying the potential factors and indicators, I furthered examined the model for potential areas of model strain or misfit by conducting a CFA on the calibration sample, which was the same dataset that I used to conduct the EFA (Byrne, 2012). I examined factor loadings, communalities, and modification indices, which allowed me to conduct post hoc analysis and address potentially problematic issues with the engagement model prior to cross-validating the model. I conducted the calibration CFA based on the model that transpired during the EFA, but I made one slight adjustment. I tested the model as a second-order factor model in which students' behavioral, emotional, and cognitive engagement with their course material represented the broader construct of academic engagement. Second-order factor models require at least three first-order factors to obtain an identifiable solution (Muthén, 2008), so I was unable to represent the social engagement constructs by a second-order factor.

Table 8

Quartimin Rotated Loadings for the Five-Factor Engagement Model from the Exploratory Factor Analysis

			Academic Engagement			Social Engagement	
Item	Item Code	Item Description	Behavioral Engagement	Emotional Engagement	Cognitive Engagement	Emotional Engagement	Cognitive Engagement
1.	q_9_8	It was easy to find and access the work that I needed to do for this course each week.	**0.897**	-0.034	0.02	-0.022	0.033
2.	q_9_7	I knew what I needed to do for this course each week.	**0.883**	0.009	0.022	-0.031	0.019
3.	q_9_9	I participated in all course assignments and activities.	**0.679**	0.106	-0.037	0.074	-0.032
4.	q_9_10	I completed all of my assignments by the due date.	**0.621**	0.001	0.005	0.124	-0.058
5.	q_10_2	The course materials and/or activities sustained my interest.	-0.015	**0.973**	-0.006	0.01	0.011
6.	q_10_1	I enjoyed the course materials and/or activities.	0.013	**0.931**	0.003	0.003	0.015
7.	q_4_2	I am very interested in the subject area of this course.	0.023	**0.662**	0.056	-0.008	-0.062
8.	q_10_5	The course materials and/or activities caused me to reflect on my understanding of the course content.	0.021	0.019	**0.931**	-0.019	0.015
9.	q_10_4	I found the course materials and/or activities to be intellectually challenging.	-0.031	-0.023	**0.849**	0.048	-0.029
10.	q_10_7	The course material and/or activities helped me understand key course concepts and facts.	0.234	0.275	**0.403**	-0.048	0.117
11.	q_8_3	I often interacted with my classmates.	0.037	-0.005	-0.006	**0.978**	-0.054
12.	q_8_1	I developed a connection with my classmates.	-0.063	0.019	0.005	**0.807**	0.099
13.	q_8_4	I enjoyed my interactions with my classmates.	0.048	0.031	0.077	**0.633**	0.179
14.	q_11_7	My interactions with classmates increased my understanding of course material.	0.052	-0.052	-0.003	-0.01	**0.93**
15.	q_11_5	My classmates made me rethink ideas that I had about course concepts.	-0.086	0.045	0.024	-0.033	**0.911**
16.	q_11_3	I learned how to interact more effectively with classmates to enhance my learning.	-0.015	-0.03	0.026	0.094	**0.826**
17.	q_11_6	I felt comfortable sharing my thoughts and opinions with my classmates.	0.109	0.041	-0.029	-0.013	**0.774**
18.	q_11_1	My classmates valued my thoughts and opinions.	0.003	0.067	-0.006	0.1	**0.744**

Note. Factors underneath social engagement represent student-student or students' engagement with their classmates; factors below academic engagement refers to students' engagement with their course content or material.
Factor loadings > .40 are in boldface.

I began this analysis by first examining the standardized factor loadings and communalities associated with each indicator. As expected, the factor loadings were all strongly related to their respective factor; none of the items were less than .60, and 15 of the items exceeded .80. Similar findings emerged when examining communalities or R^2. The three items with the lowest factor loading estimates, also contained the lowest R^2 values, but these items still contained moderately strong values above .40. Since this examination did not yield any glaring issues, I examined modification indices. Two pairs of indicators contained extremely large MIs, which both loaded onto the academic form of behavioral engagement. The first two items, which contained an MI value of 125.97, asked students to rate the extent to which they agreed with the following statements: "I participated in all course assignments and activities" (q_9_9), and "I completed all of my assignments by the due date" (q_9_10). The other two items contained an MI value of 75.01, and these items asked students about their ability to find and access their weekly course work (q_9_8), and their understanding of weekly course requisites (q_9_7). These items likely contained such large values, because these items were similarly worded (Brown, 2006). Instead of immediately correlating their error terms, I reviewed other statistics associated with these items.

After identifying items with large MI values, I returned to the factor loading and R^2 estimates. The item that asked students to indicate the extent to which they completed all of their assignments by the due date (q_9_10) had the lowest communality value and the lowest parameter estimate of all indicators in the model (i.e., .407 and .638, respectively); therefore, I deleted this item from the model and conducted another EFA. Removing this item from the model reduced the χ^2 value by nearly 180 points and resulted in the other pair of MIs becoming insignificant. The MIs revealed several other potentially problematic indicators, but implementing MIs should not be done to simply improve model fit. Instead, error terms should only be correlated if they are justified theoretically, and/or the sources of covariation are due to some outside problem that is not associated with the common factor, such as measurement, interpretation, or social disability issues (Brown, 2006). 2006).

Based on these suggestions, I decided not to implement any MIs. I did, however, remove one more item from the model: "I felt comfortable sharing my thoughts and opinions with my classmates" (q_11_6). I elected to discard this item, because it did align with my initial conceptualization of the student-student cognitive engagement nor with other items associated with this factor. Removing this item from the model reduced the overall χ^2 value by more than 97 points. These model modifications improved the overall fit of the model; however, the improvements to the χ^2 estimates were expected, since I calibrated the model on the same dataset that I used to conduct the EFA. These findings merely provided preliminary support for improving the model, so I proceeded with the analysis and attempted to validate the five constructs that emerged on the second half of the randomly split dataset.

4.6. Model Validation

After establishing the five-factor engagement measurement model through the EFA and calibrating the model, I proceeded to cross-validate the model on the second-half of the randomly split dataset. The findings indicated that the final items in the engagement measurement model, after removing the two indicators, moderately to strongly represented these constructs: $\chi^2(98) = 340.45$, $p <.001$; CFI = .97; TLI = .96; RMSEA = .065 [.058; .073]; and SRMR = .044. The RMSEA goodness-of-fit index was right at the suggested cutoff value for a strong-fitting model, but the lower-bound confidence interval was within limits, and all other model statistics strongly supported this solution. The standardized pattern coefficients revealed that the item-factor relationships were very strong, ranging from .60 to .95, and fourteen items contained standardized loadings that exceeded .75 (see Figure 9). The engagement subtypes strongly loaded onto the academic engagement construct, ranging from .61 to .92. Cognitive engagement and emotional engagement were the strongest academic engagement predictors with parameter estimates of .90 and .92, respectively.

After validating the engagement model, I conducted another CFA on all items that initially emerged from the EFA to determine if the indicators that I discarded during the model calibration analysis impacted the goodness-of-fit indices and overall model fit. Model statistics supported my decision to remove these items from the model: $\chi^2(129) = 725.18$, $p < .001$; CFI = .93; TLI = .92; RMSEA = .09 [.083; .096]; and SRMR = .05. All fit statistics markedly improved after removing these items from the model, and the χ^2 estimate reduced close to 400 points; thus, I was confident in my decision to delete these two variables. Furthermore, the relationship between items and their factors aligned with current literature. While the data provided initial support for characterizing engagement by these academic and social engagement subtypes, further analyses needed to be conducted to determine the applicability of the model to students enrolled in various course settings and to students from various ethnic and economic backgrounds.

4.6.1. Model Comparison & Multidimensionality. After validating the second-order engagement model, I examined the findings from a first-order model.

This allowed me to identify the intercorrelations between the academic engagement constructs and determine if the model should be treated as a second-order solution. As expected, the first-order solution also strongly fit the data: $\chi^2(98) = 322.63$, $p < .001$; CFI = .97; TLI = .96; RMSEA = .065 [.057; .073]; and SRMR = .04. In order to accept second-order solutions, first-order models must not be significantly different from the second-order model (Dimitrov, 2010). The $\Delta\chi^2$ suggested these models were significantly different: $\Delta\chi^2(\Delta 4) = 17.83$; $p < 0.01$. However, as previously noted, the likelihood ratio or $\Delta\chi^2$ are sensitive sample size and model complexity (Cheung & Rensvold, 2002); therefore, I conducted a ΔCFI, which indicated the models were not significantly different (ΔCFI $p < .001$). Furthermore, all other fit indices were practically identical. Thus, I concluded that I could treat the model as a second-order solution during all subsequent invariance tests.

As illustrated in Figure 10, the first-order academic engagement constructs were strongly correlated. The strength of these relationships confirmed my hypothesis regarding the multidimensional nature of the engagement constructs. The parameter estimates of the academic engagement subtypes further supported the decision to treat the model as a second-order factor model. As illustrated in Figure 9, all first-order academic engagement constructs were strongly correlated, ranging from .53 to .83. The strongest correlation was between cognitive engagement and behavioral engagement ($r = .83$, $p < .001$) (see Figure 9). The social engagement constructs were also moderately to strongly correlated with the academic engagement subtypes; however, these correlations were not as strong as the interrelationships between the academic engagement constructs. The strongest correlation was between student-student cognitive engagement and students' emotional engagement ($r = .50$, $p < .001$).

4.7. Measurement Invariance Across Course Format

The final set of covariance analyses that I performed for this dissertation aimed to determine the extent to which the recently validated measurement model functioned invariantly or similarly across the groups of students that I identified for this study. To this end, I conducted three multiple group (MG) CFAs, which each consisted of seven different invariance assessments or tests. The first step in the analysis was to ensure that model adequately fit the data across the groups being compared; thus, I started this invariance analysis by ensuring that the second-order factor model separately fit the data for both students enrolled in online courses (n = 721) and students enrolled in face-to-face courses (n = 432). The data yielded slightly better fit indices when examined among students enrolled in traditional college courses, but the model was found to adequately fit both groups of students (see Table 9). These findings confirmed that I could proceed with the next stage in the analysis and test the model for configural invariance.

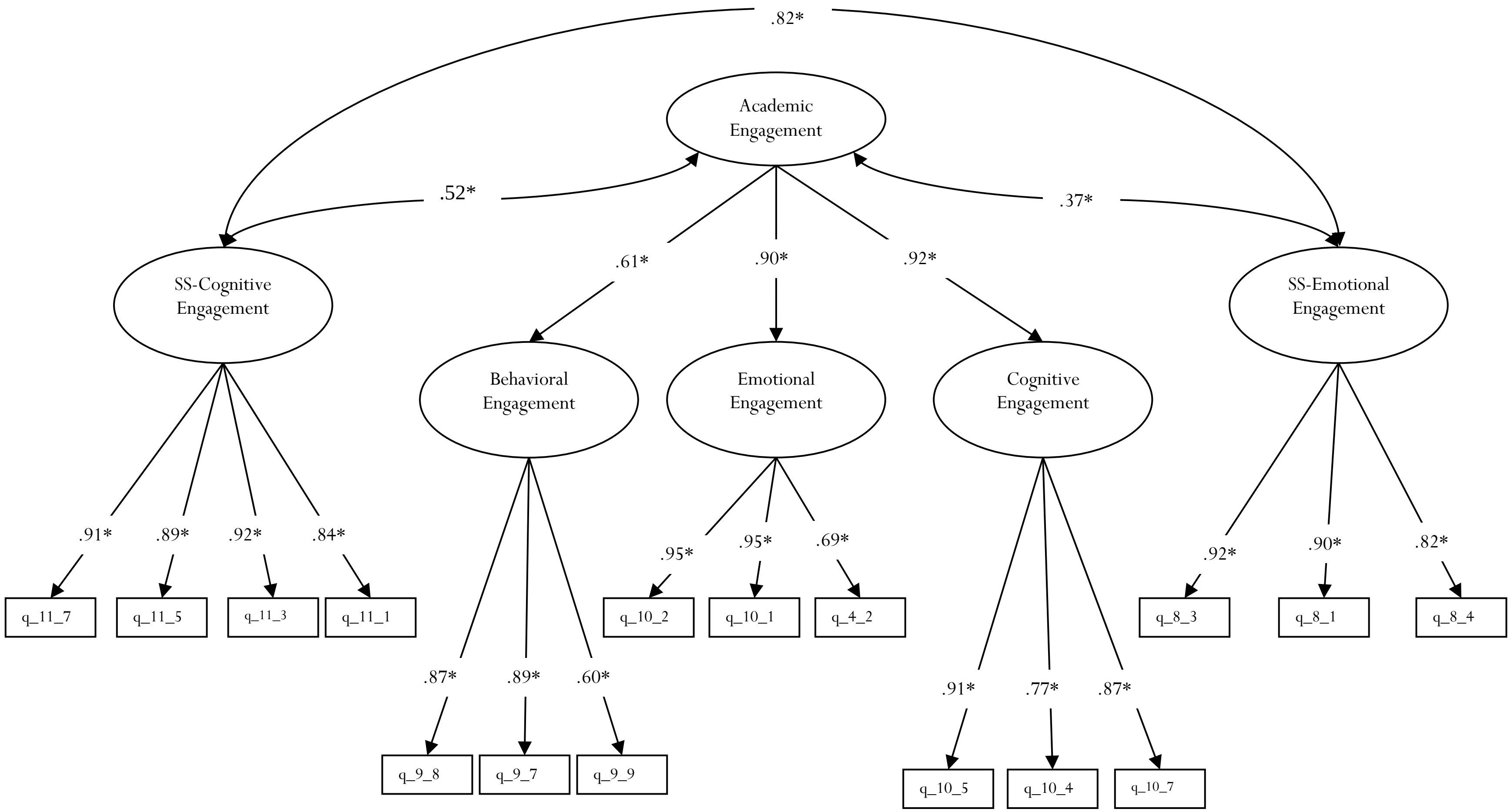

Figure 9. Restructured second-order, model of engagement with completely standardized parameter estimates to be tested for invariance across students from different course and ethnic and economic backgrounds;
SS refers to student-student;
* All parameters estimates were significant at $p < .001$.

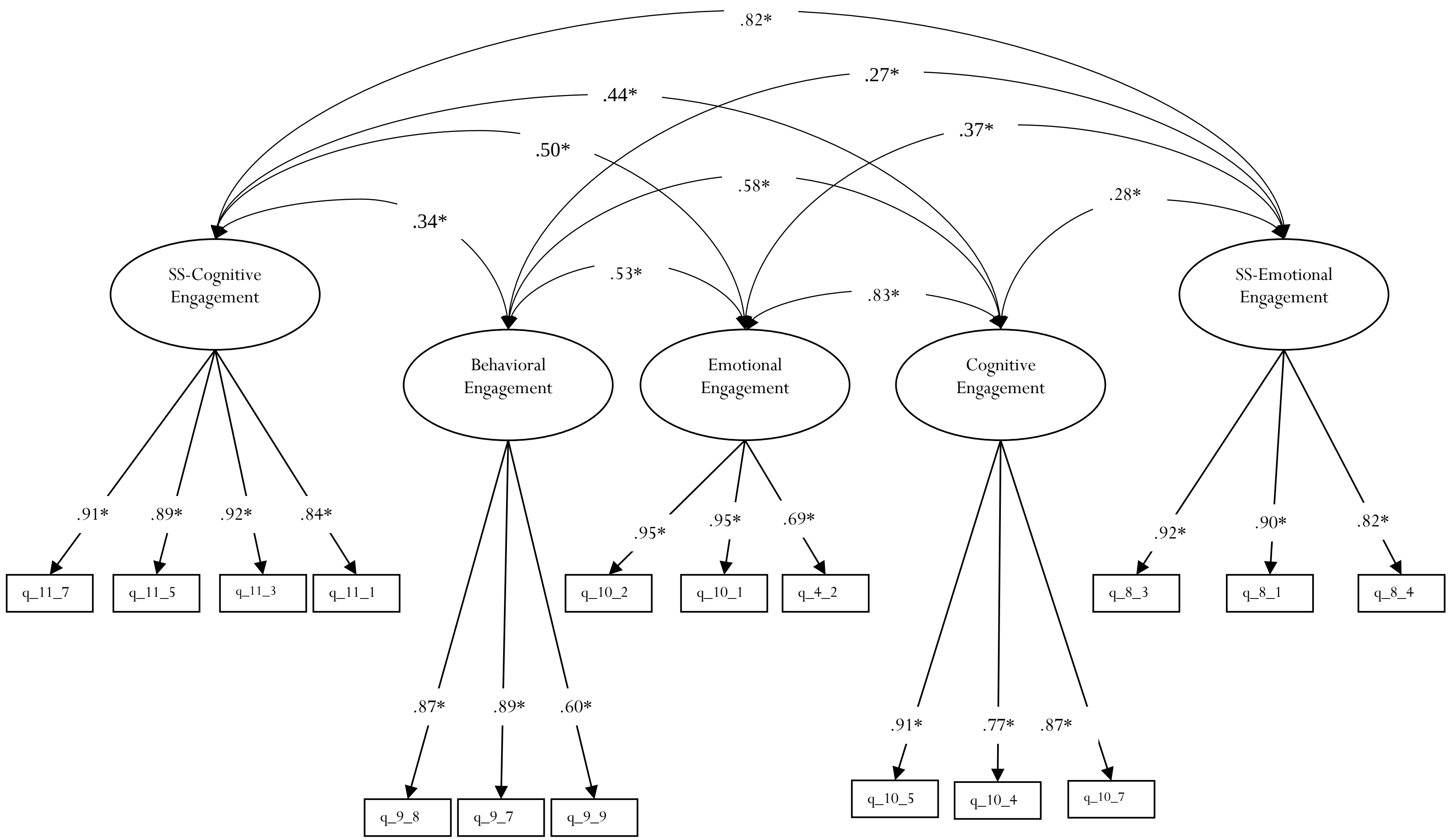

Figure 10. Path diagram of first-order engagement measurement model with correlations between all model constructs.
SS refers to student-student.
*All parameter estimates are standardized, and all items were significant at $p < .01$.

4.7.1. Configural Invariance. After determining the model adequately fit both groups of students separately, I tested the second-order engagement model for configural invariance simultaneously across students enrolled in online and in-person courses. This analysis revealed that the model adequately fit the data, which indicated that the constructs and item-factor relationship functioned similarly between these two groups of students (Brown, 2006; Byrne, 2012; Chen, et al, 2005; Cheung & Rensvold, 2002). The fit statistics for this configural model are listed in Table 9, along with the fit statistics associated with the first two analyses that I performed separately on students enrolled in online courses and students enrolled in face-to-face courses. The findings from this analysis or configural model served as the baseline for subsequent invariance tests. The item-factor relationship for the first-order social engagement constructs as well as the first-order factor-second order factor relationships between the three engagement subtypes and the higher-ordered academic engagement construct are reported for both groups of students in Figure 11. The parameter estimates indicated all items and factors were strongly related.

Table 8

Model Fit for the CFA Conducted on Each Group of Students Separately and Simultaneously for Configural Invariance

Model	$\chi 2$	*df*	CFI	TLI	RMSEA	90% CI		SRMR
Online Students	399.38	98	.967	.959	.067	[.060,	.074]	.048
In-Person Students	267.65	98	.971	.964	.063	[.054,	.072]	.039
Configural	667.03	196	.968	.961	.066	[.060,	.071]	.045

4.7.2. Metric Invariance of First-Order Factors. After establishing configural invariance, I conducted the metric invariance tests of equality of factor loadings among first and second-order factors. In order for metric invariance to be established, the factor loadings must be equivalent across both groups, which represent the strength of the linear relationship between factors in the model and the items designated to represent these factors (Chen et al., 2005). There were very few differences in model fit between the configural model and this metric model. Furthermore, the $\Delta\chi^2$ confirmed that the first-order factors functioned invariantly between groups of students: $\Delta\chi^2(\Delta 11) = 15.50$; $p = 0.16$; ΔCFI <.001. Table 10 illustrates the findings from the metric invariance test that I performed on the first-order factors.

Table 9

Model Fit from Metric Invariance Test of First-Order Factors and Chi-Square Difference Test with Configural Model

Model	$\chi2$	*df*	CFI	TLI	RMSEA	90% CI		SRMR
Metric	682.53	207	.968	.963	.064	[.059,	.070]	.046
Configural	667.03	196	.968	.961	.066	[.060,	.071]	.045
χ^2 Difference	15.50	11	.00	.02	−.002	[-.001,	-.001]	.001

4.7.3. Metric Invariance of Second-Order Factors. In addition to examining the equivalency of first-order factor loadings, I tested the second-order factor loadings for invariance to determine if these loadings significantly differed between students enrolled in online courses and students enrolled in face-to-face courses. During this test both first- and second-order factor loadings were constrained. I then compared the fit statistics of these two models (see Table 11). The $\Delta\chi^2$ and ΔCFI indicated that the second-order factor loadings functioned similarly between these groups of students: ($\Delta\chi^2(\Delta 2) = 2.31$; $p = .32$); (ΔCFI <.001). These two tests confirmed that both the first- and second-order factor loadings of this engagement measurement model functioned invariantly across these two groups of students.

Table 10

Model Fit from Scalar Invariance Test and Chi-Square Difference Test between Metric and Scalar Statistics

Model	$\chi 2$	*df*	CFI	TLI	RMSEA	90% CI		SRMR
2nd Order Metric	684.84	209	.968	.963	.064	[.059,	.069]	.048
1st Order Metric	682.53	207	.968	.963	.064	[.059,	.070]	.046
χ^2 Difference	2.31	2	.000	.000	.000	[.000,	−.001]	.002

Note. Since the $\Delta\chi^2$ and ΔCFI were non-significant, the factor loadings between these models contained comparable first- and second-order factor loadings: ($\Delta\chi^2(\Delta 2) = 2.31$; $p = .32$).

4.7.4. Scalar Invariance of Item Intercepts. I continued the invariance analysis by conducting tests of scalar of strong factorial invariance on the intercepts of the survey items. The $\Delta\chi^2$ between the model with first- and second-order factor loading constrained and this model with all item intercepts constrained was significant ($\Delta\chi^2(\Delta 13) = 65.11$; $p < .001$), which is not a desired outcome of these invariance tests; however, all other fit statistics indicated there was only a marginal change between the two models (see Table 12). Furthermore, the ΔCFI suggested that the intercepts or the origin of the scales for each item functioned invariantly (ΔCFI <.01) (Cheung & Rensvold, 2002). Since the ΔCFI indicated that there were no significant differences between the models, I determined that the intercepts of measured variables functioned invariantly between students enrolled in online courses and students enrolled in face-to-face courses, and I proceeded with the next phase of the MG CFA.

Table 11

Model Fit from Scalar Invariance Test of Item Intercepts and $\Delta\chi^2$ *between Scalar and Metric Model*

Model	χ2	*df*	CFI	TLI	RMSEA	90% CI		SRMR
Scalar (Items)	749.95	222	.964	.962	.065	[.060,	.071]	.050
2nd-Order Metric	684.84	209	.968	.963	.064	[.059,	.069]	.048
χ^2 Difference	65.11**	13	-.004.	.001	.001	[.001,	.002]	.002

Note. The $\Delta\chi^2$ yielded significant results: ($\Delta\chi^2(\Delta 13) = 65.11$; $p < .01$); however, all other fit statistics, including the ΔCFI, suggested these models functioned equivalently (ΔCFI <.001) based on recommendations made by (Cheung & Rensvold, 2002).

4.7.5. Scalar Invariance of First-Order Factors. After determining that the intercepts of measured variables were equivalent across groups, I tested the intercepts of the first-order factors a part of the broader construct of academic engagement for invariance. In first-ordered factor models, attaining this level of invariance permits factor means to be compared with confidence between groups (Milfont & Fischer, 2015). In second-order models, the loadings of first- and second-order models, the intercepts of measured variables, and the intercepts of first-order factor loadings must function invariantly prior to comparing latent mean scores across groups (Chen et al, 2005; Dimitrov. 2010). Since the two social or student-student engagement factors already achieved metric and scalar invariance, I tested the intercepts of factors that were a part of the second-order academic engagement construct.

After adding the constraints to the intercepts of the first-order factors, I compared the fit from the models again (see Table 13). Similar to the last test of invariance, the $\Delta\chi^2$ was significant: ($\Delta\chi^2(\Delta 1) = 4.5$; $p < .05$), but all other indices were nearly identical. In addition, the ΔCFI suggested that the intercepts of the second-order factor loadings functioned invariantly across students enrolled in online courses and students enrolled in face-to-face courses: (ΔCFI <.001). The invariant factor loadings and indicator intercepts suggested that the operational definition of measurement items were interpreted similarly between these groups of students (Chen et al., 2005). As a result of this test, I concluded that measurement bias or differential item functioning was not present among students enrolled in online or face-to-face courses (Dimitrov, 2010).

Table 12

Fit from Invariance Test of Factor Intercepts and Comparison with Fit from Invariance Test of Indicator Intercepts

Model	$\chi2$	*Df*	CFI	TLI	RMSEA	90% CI		SRMR
Scalar (Factors)	754.45	223	.964	.962	.065	[060,	.071]	.052
Scalar (Items)	749.95	222	.964	.962	.065	[.060,	.071]	.050
χ^2 Difference	4.5*	1	.000	.000	.000	[.000,	.000]	.002

Note. The $\Delta\chi^2$ yielded significant results ($\Delta\chi^2(\Delta 1) = 4.5$; $p < .05$); but all other fit statistics, including the ΔCFI, supported model invariance.

4.7.6. Disturbance Invariance of First-Order Factors. Typically, in first-order models, the residual invariance test, which is also known as the error invariance or full uniqueness invariance test, would follow the scalar invariance test and serve as the final assessment for measurement invariance (van de Schoot et al., 2012; Vandenberg & Lance, 2000); however, in second-order factor models, the disturbances of both first-order factors are assessed in addition to the residuals of observed variables or indicators. Therefore, I constrained the disturbances of the first-order factor loadings to be equal across both groups, while also maintaining the previous constraints imposed (i.e., factor loadings of first- and second-order factors, indicator intercepts, and first-order factor intercepts). Similar to the invariance test of first-order factor intercepts, I only constrained the disturbances of first-order factors that were a part of the broader academic engagement construct.

After adding the constraints to the disturbances of first-order factors, I compared the fit statistics associated with these models (see Table 14). Neither the $\Delta\chi^2$ or the ΔCFI were significant: $\Delta\chi^2(\Delta 3) = 3.46$; $p = .33$; $\Delta\text{CFI} < .001$. Furthermore, the TLI, RMSEA, and SRMR contained nearly identical values between models. Establishing invariance of first-order factor disturbances across these groups suggests that the first-order latent factors associated with the second-order academic engagement factor contained the same degree of measurement error between groups or simply that these first-order factors are measured similarly across groups (Cheung & Rensvold, 2002). Perhaps of most importance, establishing this form of variance invariance suggests that differences between groups on the three first-order academic engagement factors (i.e., students' behavioral, emotional, and cognitive engagement) are a result of actual differences in students' course engagement (Widamen & Reise; 1997).

Table 13

Model Fit from Invariance of First-Order Factor Disturbances and Chi-Square Difference Test with Previously Constrained Model

Model	$\chi 2$	*df*	CFI	TLI	RMSEA	90% CI		SRMR
Factor Disturbance	757.91	226	.964	.962	.065	[.060,	.070]	.054
Scalar (Factor Intercepts)	754.45	223	.964	.962	.065	[.060,	.071]	.052
X^2 Difference Test	3.46	3	.000	.000	.000	[.000,	-.001]	.002

Note. The $\Delta\chi^2$ and ΔCFI yielded non-significant results, suggesting these models did not significantly differ.

4.7.7. Residual Invariance of Indicators. The final measurement invariance test that I conducted among these students was the residual invariance test of indicators. The $\Delta\chi^2$ was found to be significant ($\Delta\chi^2(\Delta 16) = 91.01$; $p < .001$), which suggested the indicators may not have been measured identically across groups; however, I continued to examine other model statistics. Similar to other analyses in which the $\Delta\chi^2$ suggested model components were non-invariant, the other fit indices indicated otherwise, and the ΔCFI yielded was non-significant (see Table 15). Although the χ^2 difference test indicated that three aspects of the model did not function invariantly, the CFI difference test suggested otherwise. Thus, these analyses confirmed that students enrolled in online and face-to-face courses interpreted the items representing the five engagement constructs similarly.

The validation of the measurement model among these two groups of students was merely the first set of invariance tests that I performed. I applied this same process and conducted these same tests to determine if the measurement model was interpreted and functioned equivalently among the other groups of students that I selected for this dissertation. The next section presents findings from the MG CFA that I performed on students from ethnicities who have historically performed well academically (i.e., Asian and Caucasian) and students from ethnicities who have historically struggled academically (African-American and Latinos).

Table 14

Model Fit from Invariance Test of Residual Variance among Indicators and Chi-Square Difference Test with Previously Constrained Model

Model	$\chi2$	*df*	CFI	TLI	RMSEA	90% CI		SRMR
Residual (Indicators)	848.92	242	.959	.960	.067	[.062,	.072]	.062
Residual (Factors)	757.91	226	.964	.962	.065	[.060,	.070]	.054
X^2 Difference Test	91.01**	16	-.005	-.002	.002	[.002,	.002]	.008

Note. The Chi-Square Difference Test yielded significant results ($\Delta\chi^2(\Delta 16) = 91.01$; $p < .001$); but all other fit statistics, including the CFI Difference Test, supported model invariance.

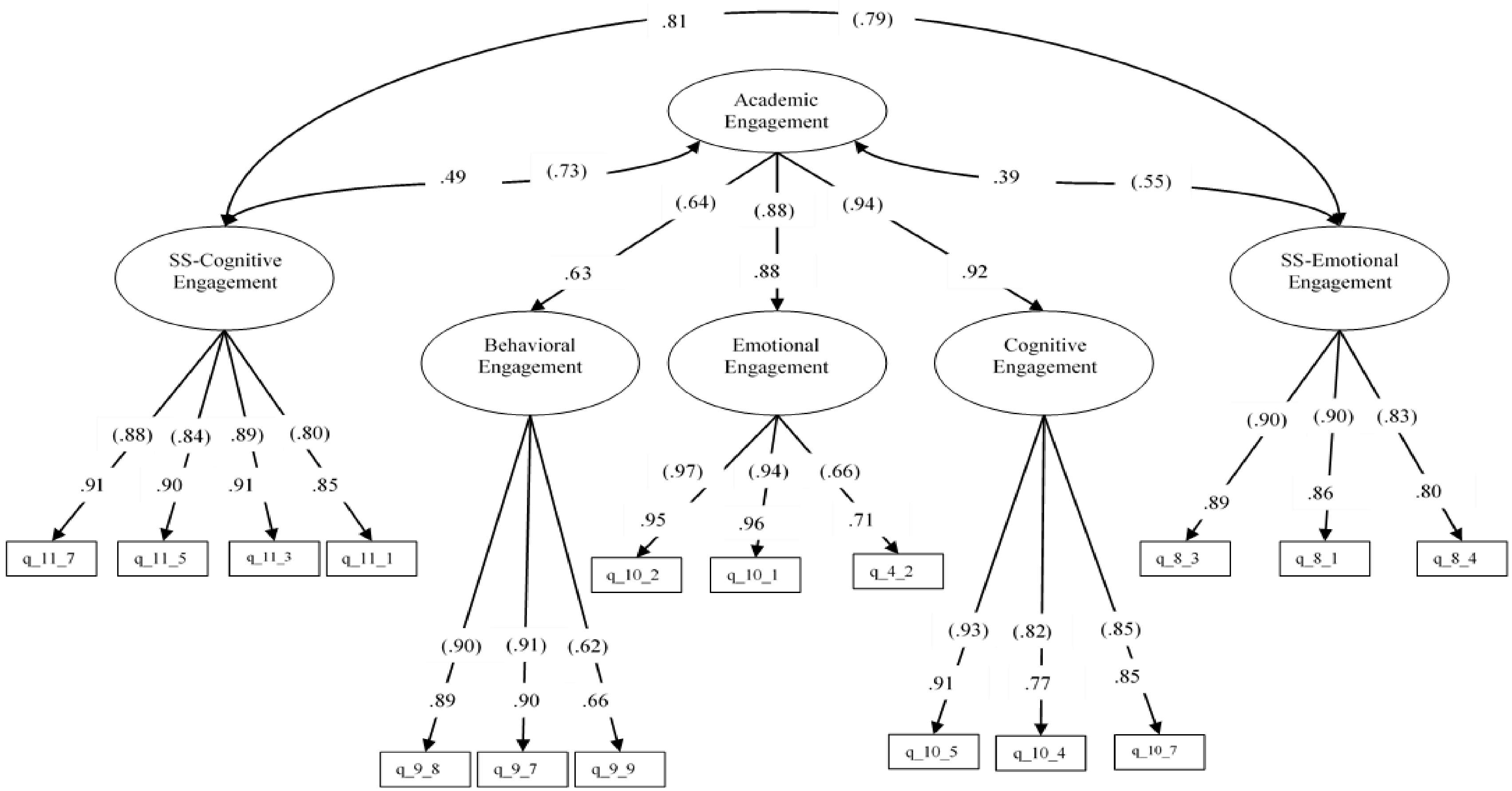

Figure 11. Results of the second-order engagement model when conducted simultaneously during configural invariance tests across students enrolled in online and in-person courses. Estimates in parentheses refer to students enrolled in face-to-face courses, while parameter estimates not parentheses refer to students enrolled in online courses.
*All parameter estimates are standardized, and all items were significant at p <.01.

4.8. Measurement Invariance across Ethnic Groups

Proceeding with the MG CFA, I tested the same model components for invariance across students who have historically performed well academically (i.e., Asian and Caucasian students) and students who have historically encountered academic difficulties (i.e., African-American and Latino students). For reporting purposes, I have labeled the merged group of students that consist of Asian and Caucasian students as "Achievers" in the tables reporting fit indices, and I have labeled the merged group of students that consists of students from African-American and Latino ethnic groups as "Underachievers". As shown in Table 18, the second-order engagement measurement model adequately fit the data when I tested it on the dataset containing only students from Asian and Caucasian ethnicities (n = 413). I obtained similar fit statistics when I tested this model on the dataset that consisted of students from African-American and Latino students (n = 288), but the model fit slightly better for these students than for students who have historically performed well academically (see Table 16). As expected, the tests of configural invariance also adequately fit the data. These tests, as well as all additional invariance tests, are reported in Table 17. The item-factor relationship for the first-order social engagement constructs as well as the first-order factor-second order factor relationships between the three engagement subtypes and the higher-ordered academic engagement construct are reported for both ethnic groups in Figure 12. The parameter estimates indicated all items and factors were strongly related, with the loadings of items and factors ranging from .63 to .95.

When I compared the fit from the configural invariance test to the fit from the metric invariance test of first-order factor loadings, the $\Delta\chi^2$ indicated that there was a significant difference between these two models ($\Delta\chi^2(\Delta 11) = 20.32$; $p < .05$). However, the model fit indices of these were quite similar to one another. The ΔCFI values indicated that these two models were not significantly different; thus, establishing metric invariance of first-ordered factor loadings. Continuing with the invariance examination of second-order factor loadings, both the $\Delta\chi^2$ and ΔCFI indicated that the loadings of the second-order factors functioned similarly across these two groups of students, since both these tests yielded non-significant findings: ($\Delta\chi^2(\Delta 2) = .31$; $p = .86$); (ΔCFI <.001). These invariance test confirmed that the factor loadings were equivalent across students who have historically performed well academically and students who have historically struggled academically

Table 15

Model Fit for the CFAs Conducted Simultaneously on all Students and on Each Group of Students Based on Ethnic Group

Model	χ2	*df*	CFI	TLI	RMSEA	90% CI		SRMR
Achievers	301.72	98	.965	.957	.071	[.062,	.080]	.051
Underachievers	231.57	98	.970	.963	.069	[.057,	.080]	.051
Configural (M0)	533.29	196	.967	.960	.070	[.063,	.077]	.051

The tests of scalar invariance also indicated that there were no significant differences between models. While the $\Delta\chi^2$ was significant ($\Delta\chi^2(\Delta 13) = 24.62$; $p < .05$), the ΔCFI suggested that the intercepts of the observed variables were invariant (ΔCFI <.001). Thus, I proceeded to test the equivalency of the first-order factor loading intercepts. The fit for this model was identical to the fit from the prior model, indicating that the models were not significantly different. Furthermore, the χ^2 critical value was not significant ($\Delta\chi^2(\Delta 13) = 4.51$; $p = .98$), and the ΔCFI further confirmed these findings: (ΔCFI <.001) (see Table 17).

After I established metric and scalar invariance across students who have historically performed well academically and students who have historically struggled, I tested the disturbance of factors and residual of indicators for invariance. The disturbances of the first-order factors were found to function equivalently based on the $\Delta\chi^2$: ($\Delta\chi^2(\Delta 3) = 4.78$; $p = .19$). The ΔCFI confirmed that this aspect of the model functioned similarly between these students (ΔCFI <.001). The final measurement invariance test that I performed was the residual invariance of indicators test. The $\Delta\chi^2$ suggested non-invariance ($\Delta\chi^2(\Delta 16) = 62.26$; $p < .01$), but the ΔCFI confirmed this aspect was invariant (ΔCFI <.001). Through the range of invariance tests that I performed, I concluded that students from ethnicities who have historically performed well academically interpreted all aspects of the engagement measurement model similar to students from ethnicities who have historically encountered academic difficulties.

4.9. Measurement Invariance across Economic Backgrounds

The last two groups of students that I compared and examined for invariance were students classified by university data as being low income and students who were not classified as being low income. As shown in Table 18, the model met accepted standards of fit across students from low-income backgrounds (n = 489) and students from higher-income backgrounds (n = 425). The configural invariance test that I performed on these groups of students also adequately fit the data. The item-factor relationship for the first-order social engagement constructs as well as the first-order factor-second order factor relationships between the three engagement subtypes and the higher-ordered academic engagement construct are reported for both economic groups in Figure 13. The parameter estimates indicated all items and factors were strongly related, with factor loadings of items and factors ranging from .60 to .96. Thus, I proceeded to conduct the subsequent invariance tests.

Table 16

All Second-Order Measurement Model Invariance Tests Performed on Historically High and Low Achieving Students with Chi-Square and CFI Difference Tests at Each Round of Invariance

Model	$\chi 2$	*df*	CFI	TLI	RMSEA	90% CI		SRMR
Configural (M0)	533.29	196	.967	.960	.070	[.063,	.077]	.051
Metric 1st-Order Factors (M1)	553.61	207	.966	.961	.069	[.062,	.076]	.054
χ^2 Difference (M1-M0)	20.32*	11	-.001	-.001	-.001	[-.001,	-.001]	.003
Metric 2nd-Order Factors (M2)	553.92	209	.966	.961	.069	[.062,	.076]	.054
χ^2 Difference (M2-M1)	.31	2	.000	.000	.000	[.000,	.000]	.000
Scalar of Indicators (M3)	578.54	222	.965	.962	.068	[.061,	.074]	.056
χ^2 Difference (M3-M2)	24.62*	13	-.001	.001	-.001	[-.001,	-.002]	.002
Scalar of Factors (M4)	583.05	223	.965	.962	.068	[.061,	.075]	.058
χ^2 Difference (M4-M3)	4.51*	1	.000	.000	.000	[.000,	.001]	.002
Residual of Factors (M5)	587.83	226	.965	.962	.068	[.061,	.074]	.062
χ^2 Difference (M5-M4)	4.78	3	.000	.000	.000	[.000,	-.001]	.004
Residual of Indicators (M6)	650.09	242	.960	.960	.069	[.063,	.076]	.066
χ^2 Difference (M6-M5)	62.26*	16	-.005	-.002	.001	[.002,	.002]	.002

Note. * = $p < .05$.

After determining that the data was suitable to proceed with invariance assessments across these two groups of students, I performed the metric invariance of first order factors test. I compared model fit from this model to the fit from the configural test or baseline model. The $\Delta\chi^2$ and ΔCFI were not significant: $\Delta\chi^2(\Delta 11) = 14.66$; $p = .20$; ΔCFI <.001. Similarly, the metric model with first and second-order factors constrained was also non-significant during the $\Delta\chi^2$ ($\Delta\chi^2(\Delta 2) = 1.66$; $p = .44$) and the ΔCFI (ΔCFI <.001). The $\Delta\chi^2$ indicated that the invariance test of indicator intercepts and the invariance test of first- and second-order factor loadings was significant ($\Delta\chi^2(\Delta 13) = 32.81$; $p < .01$); however, like all prior tests, the ΔCFI was not substantially different between these models (ΔCFI <.001), suggesting this aspect of the model was invariant.

Table 17

Model Fit for CFA Conducted Separately on Each Group of Students and on Each Group Simultaneously for the Configural Invariance Test

Model	$\chi 2$	*df*	CFI	TLI	RMSEA	90% CI		SRMR
Low-Income	304.95	98	.967	.959	.070	[.062,	.080]	.045
Not Low-Income	291.75	98	.971	.964	.064	[.055,	.072]	.045
Configural (M0)	596.70	196	.969	.962	.067	[.061,	.033]	.045

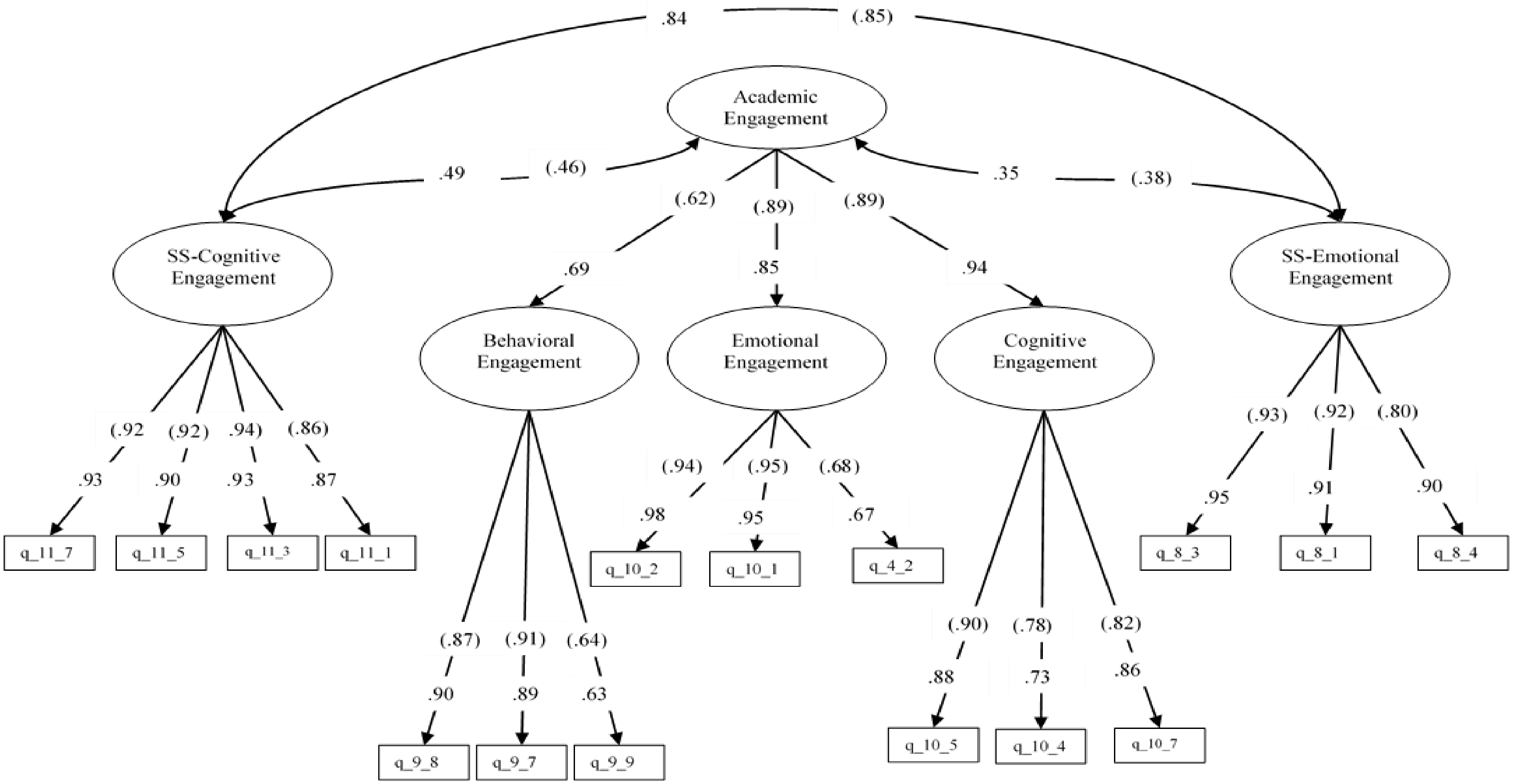

Figure 12. Results of the second-order engagement model when conducted simultaneously for configural invariance test across students classified as historically high- and low-achieving ethnic groups. Estimates in parentheses refer to students from Asian & Caucasian ethnicities, while parameter estimates not parentheses refer to students from African-American & Latino/a ethnicities.

*All parameter estimates are standardized, and all items were significant at p <.01.

The scalar test of invariance indicated that the intercepts of the observed variables were not significantly different between students from low-income backgrounds and students from higher income backgrounds ($\Delta\chi^2(\Delta 1) = .31$; $p = .58$); (ΔCFI <.001). The last two invariance tests were conducted to compare the factor disturbances ($\Delta\chi^2(\Delta 3) = 1.82$; $p = .61$); (ΔCFI <.001) and residual variances of each observed variable ($\Delta\chi^2(\Delta 16) = 20.77$; $p = .19$); (ΔCFI <.001). These model aspects also functioned invariantly across these groups, indicating that the engagement measurement model was also invariant across students from higher and lower economic statuses or backgrounds. I summarize the fit statistics for each invariance test that I performed below in Table 19.

4.10. Comparison of Latent Mean Scores

After determining that the measurement model functioned equivalently across the groups of interest, I compared students' mean scores on the final measurement model constructs. Since latent mean scores of first-order constructs are not provided when estimating second-order models, I also treated all model components as first-order constructs in order to obtain latent mean scores on the three first-order academic engagement subtype[4]. To compare latent mean scores, one of the groups in all three analyses were selected to serve as the reference group. The second-order latent mean of the reference group was fixed to zero, and the higher-ordered latent mean was allowed to freely estimate among the non-reference group. The following constraints were also imposed on models: first- and second-order

---

[4] First-order factor means in higher-order models are conditional on the second-order construct, which prevents these latent mean scores from being obtained in second-order factors solutions (Chen, Sousa, & West, 2005).

factor loadings, intercepts of measured variables, and intercepts of first-order factors (Dimitrov, 2010; Chen et al., 2005); thus, allowing the latent mean score differences among the non-reference group to be estimated. The following sections detail the differences in latent mean scores between the groups identified for this dissertation.

4.10.1. Latent Mean Comparison Across Course Format. I first obtained and compared latent mean scores on the second-order academic engagement construct and two social engagement constructs. The latent mean scores related to the two social engagement constructs were identical in both first- and second-order models estimated, since the only differences between the two models was the treatment of academic engagement constructs. I treated students enrolled in face-to-face courses as the reference group; therefore, their latent means were fixed to zero and only estimates of latent mean differences are provided for students enrolled in online courses. The findings revealed several similarities between the two groups, but more importantly, several key differences on their levels of engagement based on their latent mean scores on these constructs. The analysis revealed that students enrolled in online course scored slightly but significantly higher on the second-order academic engagement construct than students enrolled in face-to-face courses (Academic Engagement: Standardized Estimate = 0.14, p = .034). Thus, students enrolled in online courses were slightly more academically engaged than students enrolled in face-to-face courses.

Students enrolled in online courses had higher latent mean scores on the academic engagement constructs, but they scored much lower than students enrolled in face-to-face courses on both social engagement constructs. The largest discrepancy was on the student-student emotional engagement construct: (Student-

Student Engagement: Standardized Estimate = -1.00, $p < .01$). Students enrolled in online courses also scored lower on the social form of cognitive engagement than students in traditional, in-person courses, but the difference was not as large as the student-student emotional engagement construct: (Student-Student Cognitive Engagement: Standardized Estimate = -.55, $p < .01$). These findings revealed that students enrolled in online courses in this sample tended to be slightly more academically engaged than students enrolled in face-to-face courses, but they were significantly less likely to be emotionally connected and cognitively engaged with their classmates. After comparing these scores, I further compared latent mean scores of each academic engagement subtype between these two groups of students.

Across these two groups of students', latent mean scores on the behavioral engagement factor were nearly identical and not statistically different between these two groups of students (Behavioral Engagement: Standardized Estimate = .025, p = .69), suggesting no major difference in students' course participation and involvement between students. While students' latent mean scores on the emotional engagement factor were slightly higher for students enrolled in online courses than face-to-face courses (Emotional Engagement: Standardized Estimate = .12, $p = .06$), these values were also not statistically significant at $\alpha = .05$ level, but they were significant at $\alpha = .10$. While the differences between these two academic engagement constructs were negligible, slightly stronger differences emerged on students' latent mean scores on the student-content cognitive engagement factor (Cognitive Engagement: Standardized Estimate = .15, $p < .05$). These findings yielded interesting differences on latent mean scores between these two groups of students. Students enrolled in online courses reported slightly higher levels of academic engagement; particularly on the cognitive engagement construct; however, these students scored much lower on both social engagement constructs than students enrolled in traditional courses.

Table 18

All Second-Order Measurement Model Invariance Tests Performed on Students from Lower and Higher Economic Households with Findings from Chi-Square and CFI Difference Tests at Each Round of Invariance

Model	$\chi 2$	*df*	CFI	TLI	RMSEA	90% CI		SRMR
Configural (M0)	596.70	196	.969	.962	.067	[.061,	.033]	.045
Metric 1st-Order Factors (M1)	611.36	207	.968	.963	.065	[.059,	.071]	.048
χ^2 Difference (M1-M0)	14.66	11	-.001	.001	-.002	[-.002,	-.002]	.003
Metric 2nd-Order Factors (M2)	613.02	209	.969	.964	.065	[.059,	.071]	.049
χ^2 Difference (M2-M1)	1.66	2	.001	.001	.000	[.000,	.000]	.001
Scalar of Indicators (M3)	645.83	222	.967	.964	.065	[.059,	.070]	.052
χ^2 Difference (M3-M2)	32.81**	13	-.001	.001	-.001	[-.001,	-.002]	.002
Scalar of Factors (M4)	646.18	223	.967	.965	.064	[.059,	.070]	.052
χ^2 Difference (M4-M3)	.35	1	.000	.001	-.001	[.000,	.000]	.000
Residual of Factors (M5)	648.00	226	.967	.965	.064	[.058,	.070]	.052
χ^2 Difference (M5-M4)	1.82	3	.000	.000	.000	[-.001,	.000]	.000
Residual of Indicators (M6)	668.77	242	.967	.967	.062	[.057,	.068]	.054
χ^2 Difference (M6-M5)	20.77	16	.000	.002	-.002	[-.001,	-.002]	.002

Note. ** = $p < .01$.

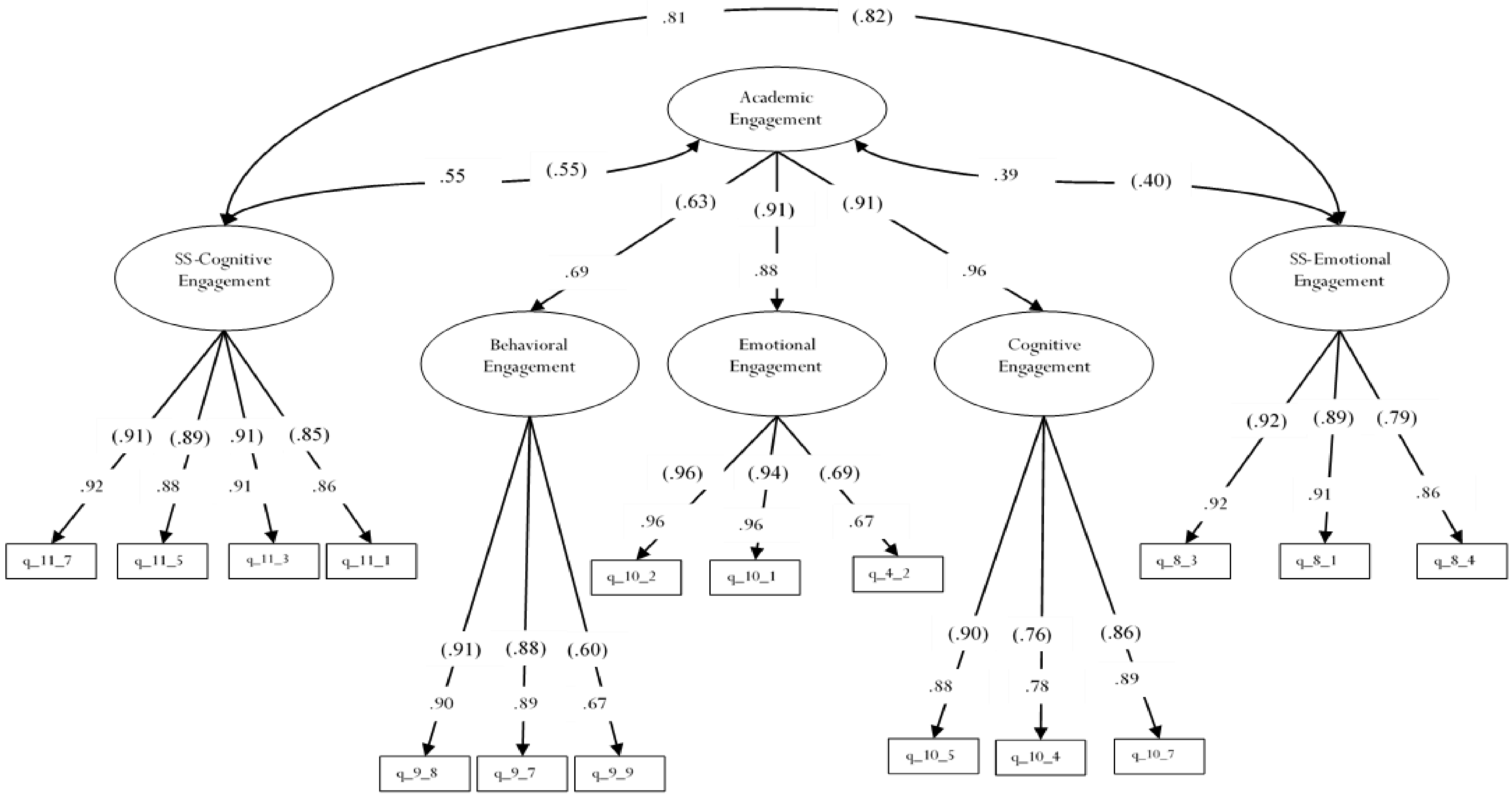

Figure 13. Results of the second-order engagement model when conducted simultaneously during configural invariance tests across students classified as being low income and students not classified as being low income. Estimates in parentheses refer to students who are not low-income, while parameter estimates not parentheses refer to students who are low-income.

*All parameter estimates are standardized, and all items were significant at p <.01.

4.10.2. Latent Mean Comparison Across Ethnic Groups. Conducting the same comparison across students who have historically performed well academically (i.e., Asian and Caucasian students) and students who have historically struggled academically (i.e., African-American and Latino/a students) resulted in similar academic and social engagement trends. During this analysis, students from ethnicities who have historically performed well academically served as the reference group, so only estimates for students from African-American and Latino/a ethnic groups are provided. Students from African-American and Latino/a ethnicities reported slightly higher scores on both academic and social engagement model constructs (Academic Engagement: Standardized Estimate = 0.18, $p = < .05$). While slight differences were found on the broader academic construct, stronger differences on the two social engagement constructs emerged. Students from ethnicities who have historically struggled academically scored moderately higher on items measuring student-student emotional engagement (Student-Student Emotional Engagement: Standardized Estimate = .31, $p < .01$) and student-student cognitive engagement (Student-Student Cognitive Engagement: Standardized Estimate = .34, $p < .01$). Thus, students who have historically encountered academic difficulties were more academically engaged with the course material and socially engaged with their classmates than students from ethnicities who have historically performed well academically.

I continued this analysis by drilling down and examining latent mean differences on the academic engagement subtypes between these two groups of students. There were no significant latent mean differences on the behavioral engagement scores (Behavioral Engagement: Standardized Estimate = .13, $p = .13$).

The emotional engagement construct was only significant at $\alpha = .10$: Emotional Engagement: Standardized Estimate = .13, $p = .09$. On both of these measures, students from ethnicities who have historically encountered academic difficulties scored slightly higher than their counterparts. While differences on the academic form of behavioral and emotional engagement were negligible, students from ethnicities who have historically struggled academically reported higher levels of cognitive engagement with their course material than students from ethnicities who have historically performed well academically (Cognitive Engagement: Standardized Estimate = .18, $p < .05$). Overall, students from African-American and Latino/a ethnicities were more academically engaged, which primarily stem from their cognitive engagement with their material. They were also more emotionally and cognitively engaged with their classmates than students from Asian and Caucasian ethnicities.

4.10.3. Latent Mean Comparison Across Economic Status. I concluded the latent mean score analysis by comparing students who were classified by institutional data as being low income to students who were not classified as being low income. Unlike the two previous analyses, students from high- and low-income backgrounds reported nearly identical scores on all model constructs. During this analysis, students from higher-income backgrounds served as the reference group, so only estimates of students who were classified as being low income were estimated. No significant differences emerged on any of the academic engagement constructs. These students also reported nearly identical scores on the social form of cognitive engagement (Student-Student Cognitive Engagement: Standardized Estimate = .096, $p = .16$). The only factor that was close to containing a significantly different mean score was the student-student emotional engagement construct (Student-Student Emotional Engagement: Standardized Estimate = .13, p = .054). Students classified as being low income reported slightly higher levels of emotional engagement with their classmates than students who were not classified as being low income. Thus, students identified by administrative data as being low income encountered similar academic engagement experiences and were slightly more emotionally engaged with their classmates than were students not classified as being low income.

5.0. Discussion & Conclusion

5.1. Findings & Implications

The recent surge in online courses among institutions of higher education and subpar performance among students in these courses, warrants scrutinizing areas that have been positively associated with student learning, development, and success in traditional course settings and determining the extent to which these aspects apply to students enrolled in online course settings. Student engagement is an area that has consistently been found to increase students' academic performance, across all levels of education (Fredericks, et al., 2004; Kuh et al., 2006; Newmann, et al., 1992). However, prior to examining the relationship between student engagement and course related outcomes, sound models capable of measuring these areas must first be established. Thus, a primary goal of the current study was to determine whether Fredericks and colleagues' (2004) conceptualization of engagement could be applied to students enrolled various college course settings, particularly online courses—given the dearth of validated measures currently in use in higher educational settings this work is of critical importance. I also wanted to provide a more comprehensive assessment of student engagement, which influenced my decision to examine engagement between students and the various types of interactions they encounter in college courses, focusing on course-level behaviors that could be used to assess the design and implementation of higher education courses, including online courses in higher education.

In the current study, I adopted Fredericks Blumenfeld, and Paris' (2004) tripartite engagement framework that characterized engagement as containing a behavioral, an emotional, and a cognitive component; adapted their model by incorporating literature on distance education and student interactions; apply the three engagement subtypes to the interactions students typically encounter in college courses, including their interactions with their course material, with their classmates, and with their instructors/teaching assistants; and assessed the validity of the adapted model across students enrolled in different courses and across students from different ethnic and economic backgrounds. The model that I proposed and tested partitioned engagement into academic and social domains by treating the academic and social engagement subtypes as first-order constructs represented by these two higher-ordered factors. I proposed such a comprehensive model of engagement to deconstruct specific sources of engagement, which would allow me to identify how and with whom students were engaging. The 12-factor, second-order measurement model of engagement that I proposed and tested did not meet accepted standards of fit. Thus, my initial hypothesis that the proposed engagement measurement model would meet accepted standards of fit was rejected, and I explored alternative model solutions using a combined exploratory and confirmatory factor analytic approach.

The data used to test the initial engagement model did not support examining engagement between students and their instructors/TAs, but I was able to address my second research question and cross-validate an alternative engagement model. The final model that emerged contained both academic and social engagement constructs. The final five-factor solution consisted of students' behavioral, emotional, and cognitive engagement with their course material and students' emotional and cognitive engagement with their classmates. The constructs in the final engagement measure model were strongly supported by the data, and all items moderately to strongly loaded onto their respective factor. I was also able to confirm the multidimensionality of the model by comparing fit from the second-order solution to a factor structure in which all latent variables were treated as first-order constructs. The first-order academic engagement factors were all strongly interrelated, which confirmed my hypothesis that the engagement subtypes would be strongly correlated and represent a higher-ordered construct[5]. These findings offer promise for characterizing and measuring academic forms of engagement through behavioral, emotional, and cognitive engagement subtypes as synthesized by Fredericks, Blumenfeld, and Paris (2004). While the tripartite engagement model has primarily been used to measure engagement at the primary and secondary education level, the current study suggests that these forms of engagement appear to be suitable for examining engagement at the post-secondary education level.

The tests of invariance indicated that the model is suitable for examining engagement among a range of students in the current study sample. More specifically, the model functioned invariantly across course settings and across students from varying ethnic and economic backgrounds. This is particularly promising, since there does not appear to be a universally accepted approach for measuring engagement in online course settings. As previously noted, researchers often modify existing instruments or create their own scales to measure engagement (Dixon, 2010; Robinson & Hullinger, 2006). Furthermore, the experiences of students from minority groups and lower-income backgrounds have not been well documented in online courses. Providing students with meaningful academic

[5] Since only two social engagement constructs emerged, these factors could not be included in a second-order solution.

communities and social support systems and enhancing students' perceptions of belonging have commonly been found to foster the academic success and retention of ethnic and racial minorities and students from low-income backgrounds (Johnson, Wasserman, Yildirim, & Yonai, 2014; Oseguera, Locks, and Vega). Since both academic and social engagement model components were validated and functioned invariantly across students enrolled in online and face-to-face courses, students from differing ethnic groups, and students from lower- and higher economic backgrounds, the current model has the potential to identify the extent to which these students are academically and socially engaged in their course.

Identifying students' levels of academic and social engagement will allow instructors and other educators to either confirm engaging practices or restructure course activities to enhance students' engagement. As recently noted, social aspects are particularly important for students from minority and low-income backgrounds; however, these benefits are not limited to students from these backgrounds. Methods that have been found to promote the success of underrepresented low-income students are likely to be successful to the larger body of students attending college; however, designing strategies solely for the general student population without considering the specific circumstances and needs of low-income and underrepresented minority students are typically not helpful for these students (Thayer, 200). Therefore, the proposed model has the potential to benefit college instructors, course designers, and other college personnel working to enhance students' course experiences and engagement. The current model not only provides promise for uncovering levels of engagement, it allowed levels of student engagement to be identified and compared among students in the current study sample.

The findings from the latent mean comparison revealed that certain groups of students reported higher forms of academic and social engagement than others. Students enrolled online courses and students from African-American and Latino/a ethnic groups scored slightly higher on items measuring academic engagement. While these findings were significant, the actual mean difference was quite marginal in relation to their comparison groups. The most notable differences emerged on the two social engagement constructs. Technological advances allow students to interact synchronously and asynchronously, but it does not appear that technology was utilized in the courses from which the sample was drawn in ways that fostered student-student connections or higher-ordered thinking, since students enrolled in online courses scored significantly lower on the social engagement measures than students enrolled in face-to-face courses. Students enrolled online courses were, however, more cognitively engaged than their comparison groups, which may largely be due to the independent nature of these courses.

Students from African-American and Latino/a ethnicities also scored higher on both social engagement constructs than students from Asian and Caucasian ethnicities, which should enhance their college experiences and academic performance. A number of studies conducted on both African-American and Latino/a students have noted the importance of developing social connections. For instance, Oseguera and colleagues (2008) conducted a meta-analysis on student success and student retention methods; Latino/a students who developed a sense of community were more likely to persist and succeed academically, and it has been argued that students' college experiences are a stronger predictor of college adjustment and persistence than student background characteristics (Carter, 2006; Oseguera et al., 2008). Strayhorn (2008) studied the relationship between students' academic achievement, college satisfaction, and the supportive relationships that African-American students develop with their peers, with their instructors, and with college staff. African-American males who developed supportive relationships with various school members reported significantly higher levels of college satisfaction. Similarly, Johnson and colleagues (2014) found pleasant academic interactions and positive social environments to positively predict student persistence among Caucasian students and students from varying ethnic identities, including students who identified as being Latino, African-American, Asian-American, or Multiracial. Unfortunately, minority students, including Latino and African-American students, are more likely to report experiences of hostile campus environments (Hurtado & Carter, 1996). Although faculty members are unable to address every college incident or all feelings of college exclusion, they are capable of promoting an inclusive course environment. At least among students in the current study, it appears that African-American and Latino/a students were able to develop social bonds or connections with their peers that also stimulated higher-ordered thinking, since these students reported significantly higher levels of student-student emotional and cognitive engagement. Interestingly, no differences on any of the academic or social engagement constructs emerged among students classified as being low income and students not classified as being low income.

5.2. Study Limitations

The validation of the model that emerged in the current study shows promise for measuring academic and social engagement across a range of students, but there were several limitations of the study that must be addressed. One of the limitations in the study pertained to style of data collected. All of the analyses conducted was based on students' self-reported survey responses, which pose several concerns regarding the validity of the data. First, the survey was quite demanding in length. The engagement data collected was a part of broader evaluation of an online course development initiative, so there were a number of different areas that were assessed. This resulted in creating a lengthy survey to address areas of interest. Survey fatigue is a concern in long surveys. Incorporating checks within the survey to ensure students are thoroughly answering questions should be considered in future iterations of survey implementations. In addition, collecting different types of data, such as instructor interviews or surveys, may be incorporated to triangulate student engagement data and further validate student responses.

During the process of creating survey items that were believed to represent students' engagement with the course material, a check for validity was implemented. Face validity was incorporated by consulting an expert in the domain of student motivation and engagement. Ideally, multiple experts would be consulted during assessments of face validity; however, since time was a restriction, only one expert was consulted to conduct the face validity assessment. Despite this limitation, the expert believed the items seemed to adequately represent aspects of student engagement as characterized by Fredericks and colleagues (2004). This confirmation was further validated as the three academic engagement aspects, indeed, transpired in the final model.

Another limitation of the study pertained to the data available. The sample used to conduct the invariance tests consisted of some of the same students used validate the model. Ideally, the tests of invariance would have been conducted on entirely independent samples; unfortunately, sample size restrictions prevented this from happening. In order to determine whether the findings were an artifact of the data, I also divided the second half of the randomly split dataset and created datasets consisting of only students enrolled in online courses and students enrolled in face-to-face courses. I then tested the final engagement model that transpired from the EFA on both of these groups of students separately. The model did meet accepted standards of fit. I also used these datasets to test for invariance, and the same findings emerged. These findings increased my confidence in findings of the current study; thus, the current study shows promise for an instrument and model that may have great utility in measuring academic and social forms of engagement.

5.3. Future Directions

Prior research has indicated that students who demonstrate higher levels of engagement are more likely succeed and persists in their academic studies. Future studies should examine the relationship between specific forms of engagement that emerged in this model and course outcomes, such as academic performance, course satisfaction, and course success. This would allow researchers to determine whether these model constructs that emerged in this study indeed course outcomes positively. Furthermore, identifying the relationship between specific forms of engagement and college course outcomes would allow targeted recommendations that are potentially capable of increasing student learning development and success to be developed for different groups of students. Furthermore, additional attempts should be made to validate constructs related to pedagogy. Understanding drivers or predictors of academic and social forms of engagement may provide information even more useful for higher education professionals as they proceed with online course and online program development, and this information should be equally important for instructors of any college course attempting to promote engagement and success among their students.

Course instructors and designers through strategic course planning should be able to help facilitate student engagement with the course content and various course actors. Online courses require different needs from course designers and instructors to ensure students interact and engage with other course members. The lower levels of engagement among students enrolled in online courses in this study, highlights the need to identify ways to enhance students' interactions with other course members. Ensuring technology is effectively utilized to meet student needs is another aspect of online classes that must be considered when designing online courses and developing course components and activities. While technology is indeed a critical component of online classes, the current model does not incorporate any technological assessments. It seems highly plausible that negative technological experiences could result in negative perceptions or experiences students have towards the course and other course members. Future research on student engagement in online courses would benefit by accounting for the influence of technology on students' level of engagement. These findings may illuminate bright spots or effective ways technology may be incorporated to enhance student engagement.

There is a lot of variability in all college courses. This variability makes it difficult to develop standard approaches to college course engagement. Course design and implementation decisions vary from course to course. Some instructors may focus more on group or collaborative work, while others may focus more on independent work. As such, information obtained relating to each aspect of engagement may be more useful by comparing it to the instructor's intended goals for the course. If they desire to create emotional bonds and connections between students and the survey data indicates this is not occurring, this information could be relayed to the course instructors and designers to help them rethink their course structure. Furthermore, engagement may vary by discipline or content. Certain disciplines may demand more independent work, while others may require more collaboration. Thus, determining whether engagement differs by discipline is another area to pursue.

5.4. Summary & Significance

The findings from the current study revealed that a second-order model of engagement, which was primarily based on Fredericks, Blumenfeld, and Paris' (2004) multidimensional engagement framework, has implications for examining student engagement across multiple college course settings and across ethnically and economically diverse groups of students. Furthermore, the model validated in this study provides support for characterizing engagement through academic and social components or forms since the final model constructs that were validated consisted of students behavioral, emotional, and cognitive engagement with their course material and students' emotional and cognitive engagement with their classmates. One of the benefits of the final five-factor model is its potential in identifying forms and sources of engagement. This may be particularly beneficial to instructors and designers of online courses, since the quality of student interactions is an existing concern of online instruction. Instructors and/or course designers may be able to utilize information from these measures to improve or validate existing instructional practices and course components. The absence of widely used models for examining engagement in online courses provides additional significance to the current study and model.

Another benefit of the current model is its robustness and applicability to students from differing backgrounds in the study sample. The experiences of underrepresented students and students from low-income backgrounds in online courses is an area that is need of greater scrutiny. Since the current model functioned invariantly across these groups of students, the current model provides one approach for investigating the course experiences and levels of engagement among these groups of students. Among the current study sample, glaring differences emerged between students' levels of academic and social course engagement. The latent mean comparison between groups enrolled in different course formats suggested that students enrolled in online courses and students from African-American and Latino/a ethnicities were slightly more academically engaged than their counterparts. However, students enrolled in online courses scored much lower than students enrolled in face-to-face courses on the social engagement measures, while students from African-American and Latino/a ethnic groups scored higher on the social engagement measures than did students from Asian and Caucasian ethnicities. Interestingly, no differences emerged between groups of students from lower and higher economic backgrounds.

There are various potential reasons for these differences. The lack of physical space shared among students in online courses may be one reason for the lower levels of social engagement among these students, while cultural differences regarding the importance of social connectedness and community may be potential causes for increased levels of social engagement among students from African-American and Latino/a ethnicities. While the exact causes for the variation in levels of academic and social engagement cannot be identified in the current study, the findings revealed interesting differences and potential areas of further exploration. Furthermore, the model's applicability to multiple college course contexts and groups of students provides potential for great use. The current model may be utilized by college instructors, course designers, and personnel to determine the extent to which students are engaging academically and whether students are engaging in their course as expected, providing these college personnel with validation of current course design and pedagogical practices or suggestions for course structure, practices, and improvement.

6.0 References

Allen, I. E., & Seaman, J. S. (2015). Grade level: Tracking online education in the United States. *Babson Survey Research Group*.

Angelino, L. M., Williams, F. K., & Natvig, D. (2007). Strategies to engage online students and reduce attrition rates. *The Journal of Educators, 4*(2), 1-14.

Aragon, S. R. (2003). Creating social presence in online environments. *New Directions for adult and continuing education, 2003*(100), 57-68. doi: 10.1002/ace.119

Asparouhov, T. & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling, 16*(3), 397-438. doi: 10.1080/10705510903008204

Astin, A. W. (1984). Student involvement. A development theory for higher education. *Journal of College Student Personnel, 25*(4), 297-308.

Baker, C. (2010). The impact of instructor immediacy and presence for online student affective learning, cognition, and motivation. *The Journal of Educators Online, 7*(1), 1-30.

Bean, J., & Eaton, S. B. (2001). The psychology underlying successful retention practices. *Journal of*

College Student Retention, 21(1), 73-89.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88*(3), 588-606.

Bernard, R. M., Abrami, P. C., Borokhovski, E. C., Wade, A., Tamim, R. M., Surkes, M. A., & Bethel, E. C. (2009). A meta-analysis of three types of interaction treatments in distance education. *Review of Educational Research, 79*(3), 1243-1289. doi: 10.3102/0034654309333844

Brown, T. A. (2006). *Confirmatory factor analysis for applied research.* New York: Guilford.

Bullen, M. (1998). Participation and critical thinking in online university distance education. *Journal of Distance Education, 13*(2), 1-32.

Burdenski, T. (2000). Evaluating univariate, bivariate, and multivariate normality using graphical and statistical procedures. *Multiple Linear Regression Viewpoints, 26*(2), 15-28.

Byrne, B. M. (2012). *Structural equation modeling with Mplus: Basic concepts, applications, and programming.* New York, NY. Taylor and Francis Group.

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological bulletin*, *105*(3), 456.

Byrne, B. M., & Watkins, D. (2003). The issue of measurement invariance revisited. *Journal of Cross-Cultural Psychology*, *34*(2), 155-175.

Carr, S. (2000). As distance education comes of age, the challenge is keeping the students. *The Chronicle of Higher Education, 46*(23), 39-41.

Carter, D. F. (2006). Key issues in the persistence of underrepresented minority students. *New Directions for Institutional Research, 2006*(130), 33-46. doi: 10.1002/ir.178

Chen, F. F., Sousa, K. H., & West, S. G. (2005). Testing measurement invariance of second-order factor models. *Structural Equation Modeling, 12*(3), 471-492.

Chen, P. S. D., Lambert, A. D., & Guidry, K. R. (2010). Engaging online learners. The impact of Web based learning technology on college student engagement. *Computers & Education, 54*(4), 1222-1232. doi:10.1016/j.compedu.2009.11.008

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural equation modeling*,*9*(2), 233-255.

Chickering, A. W., & Gamson, Z. F. (1987). Seven principles for good practice in undergraduate education. *American Association for Higher Education Bulletin,* 39(7), 3-7.

Conway J. M., & Huffcutt A. I. 2003. A review and evaluation of exploratory factor analysis practices in organizational research. *Organizational Research Methods*, 6(2)147-168. doi: 10.1177/1094428103251541

Corno, L., & Madinach, E. B. (1983). The role of cognitive engagement in classroom learning and motivation. *Educational Psychologist, 18*(2), 88-108.

Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation, 10*(7), 1-8. Retrieved from http://pareonline.net/pdf/v10n7.pdf

Dai, D. Y., & Sternberg, R. J. (2004). Beyond Cognitivism: Toward an integrated understanding of intellectual functioning and development. New Jersey: Lawrence Erlbaum Associates.

Dimitrov, D. M. (2006). Comparing groups on latent variables: A structural equation modeling approach. *Work*, *26*(4), 429-436.

Dimitrov, D. M. (2010). Testing for factorial invariance in the context of construct validation. *Measurement and Evaluation in Counseling and Development*, *43*(2), 121-149.

Dixson, M. D. (2010). Creating effective student engagement in online courses: What do students find engaging? *Journal of the Scholarship of Teaching and Learning, 10*(2), 1-13.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4*(3), 272-299.

Ferguson, E. & Cox, T. (1993). Exploratory factor analysis: A user's guide. *International Journal of Selection and Assessment, 1*(2), 84-94.

Finn, J. D. (1989). Withdrawing from school. *Review of Educational Research, 59*(2), 117-142. doi: 10.3102/00346543059002117

Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment, 7*(3), 286-299.

Ford, J. K., MacCallum, R. C., Tait, M. (1986). The application of exploratory factor analysis in applied psychology: A critical review and analysis. *Personnel Psychology, 39*(2), 291-314.

Fredericks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research, 74*(1), 59-109. doi: 10.3102/00346543074001059

Garrison, D. R., Anderson, T., & Archer, W. (2000). Critical inquiry in a text-based environment: Computer conferencing in higher education. *The Internet and Higher Education, 2*(2-3), 87-105.

Garrison, D. R., Anderson, T., & Archer, W. (2001). Critical thinking and computer conferencing: A model and tool to access cognitive presence. *American Journal of Distance Education, 15*(1), 7–23.

Garrison, D. R., & Cleveland-Innes, M. (2005). Facilitating cognitive presence in online learning: Interaction is not enough. *The American Journal of Distance Education, 19*(3), 133-148.

Gokhale, A. A. (1995). Collaborative learning enhances critical thinking. *Journal of Technology Education, 7*(1), 22–30.

Goodenow, C. (1992, April). *School motivation, engagement, and sense of belonging among urban adolescent students.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Hair, J. F., Ringle, C. M., & Sarstedt, M. (2011). PLS-SEM: Indeed a silver bullet. *Journal of Marketing Theory and Practice, 19*(2), 139-151. doi: 10.2753/MTP1069-6679190202

Harrington, D. (2009). *Confirmatory factor analysis*. Oxford University Press, USA.

Hayton, J. C., Allen, D. G., & Scarpello, V. (2004). Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organizational research methods*, *7*(2), 191-205.

Henson, R. K., & Roberts, J. K. (2006). Use of exploratory factor analysis in published research: Common errors and some comment on improved practice. *Educational and Psychological Measurement, 66*(3), 393-416. doi: 10.1177/0013164405282485

Hirschfeld, G., & von Brachel, R. (2014). Multiple-group confirmatory factor analysis in R – A tutorial in measurement invariance with continuous and ordinal indicators. *Practical Assessment, Research, & Evaluation*, *19*(7), 1-12.

Hooper, D., Coughlan, J., & Mullen, M. (2008). Structural equation modeling: Guidelines for determining model fit. *Electronic Journal of Business Research Methods, 6*(1), 53-60.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1-55.

Hurtado, S. & Carter, D. F. (1997). Effects of college transition and perceptions of the campus racial climate on Latino college students' sense of belonging. *Sociology of Education, 70*(4), 324-345.

Jaggars, S. J., & Bailey, T. (2010). Effectiveness of fully online courses for college students: Response to a Department of Education meta-analysis. *Community College Research Center: Teachers College Columbia University.*

Jimmerson, S. R., Campos, E., & Greif, J. L. (2003). Toward an understanding of definitions and measures of school engagement and related terms. *The California School Psychologist, 8*(1), 7-27.

Johnson, D. R., Wasserman, T. H., Yildirim, B. A., & Yonai, B. A., (2014). Examining the effects of stress and campus climate on the persistence of students of color and white students: An application of Bean and Eaton's psychological model of retention. *Research Higher Education, 55*(1) 75-100. doi: 10.1007/s11162-013-9304-9

Johnson, H., Mejia, M. C., & Cook, K. (2015). Successful online courses in California's community colleges. *Public Policy Institute of California,* 1-24.

Kahn, J. H. (2006). Factor analysis in counseling psychology research, training, and practice: Principles, advances, and applications. *The Counseling Psychologist, 34*(5), 684-718. doi: 10.1177/0011000006286347

Kena, G., Musu-Gillette, L., Robinson, J., Wang, X., Rathbun, A., Zhang, J., Wilkinson-Flicker, S., Barmer, A., & Dunlop Velez, E. (2015). *The Condition of Education 2015* (NCES 2015-144). U.S. Department of Education, National Center for Education Statistics. Washington, DC. Retrieved from http://nces.ed.gov/pubsearch.

Komarraju, M., Musulkin, S., & Bhattacharya, G. (2010). Role of student-faculty interactions in developing college students' academic self-concept, motivation, and achievement. *The Johns Hopkins University Press, 51(*3), 332-342. doi:10.1353/csd.0.0137

Kreijns, K., Kirschner, P. A., & Jochems, W. (2003). Identifying the pitfalls for social interactions in computer-supported collaborative learning environments: A review of the research. *Computers in Human Behavior, 19*(1), 335-353.

Kuh, G. D. (2001). Assessing what really matters to student learning: Inside the National Survey of Student Engagement. *Taylor & Francis Group, 33*(3), 10-17, 66.

Kuh, G. D. (2009). What student affairs professionals need to know about student engagement. *Journal of College Student Development, 50*(6), 683-706. doi: 10.1353/csd.0.0099

Kuh, G. D., Cruce, T. M., Shoup, R., Kinzie, J., & Gonyea, R. M. (2008). Unmasking the effects of student engagement on first-year college grades and persistence. *The Journal of Higher Education, 79*(5), 540-563.

Kuh, G. D., Kenzie, J., Buckley, J. A., Bridges, B. K., & Hayek, J. C. (2006). What matters to student success: A review of the literature. *National Postsecondary Education Cooperative.*

Lamborn, S., Newmann, F., & Wehlage, G. (1992). The significance and sources of student engagement. *Student engagement and achievement in American secondary schools*, 11-39.

LaPointe, D. K., & Gunawardena, C. N. (2004). Developing, testing and refining of a model to understand the relationship between peer interaction and learning outcomes in a computer-mediated conferencing. *Distance Education, 25*(1), 83-106. doi: 10.1080/0158791042000212477

Lundberg, C. A., & Schreiner, L. A. (2004). Quality and frequency of faculty-student interaction as predictors of learning: An analysis by student race/ethnicity. *Journal of College Student Development, 45*(5), 549-565.

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods, 4*(1), 84-99.

Marks, H. M. (2000). Student engagement in instructional activity: Patterns in elementary, middle, and high school years. *American Educational Research Association, 37*(1), 153-184.

Marsh, H. W., Balla, J. R., & McDonald, R. P. (1988). Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. *Psychological bulletin*, *103*(3), 391.

Marsh, H. W., Hau, K. T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural equation modeling*, *11*(3), 320-341.

McKinney, J. P., McKinney, K. G., Franiuk, R., & Schweitzer, J. (2006). The college classroom as a community: Impact on student attitudes and learning. *College Teaching*, *54*(3), 281-284.

Meece, J. L., Blumenfeld, P. C., & Hoyle, R. H. (1988). Students' goal orientations and cognitive engagement in classroom activities. *Journal of Educational Psychology, 80*(4), 514-523.

Milfont, T. L., & Fischer, R. (2015). Testing measurement invariance across groups: Applications in cross-cultural research. *International Journal of Psychological Research*, *3*(1), 111-130.

Muthén, L. K. (2008, December 30). Re: Second order factors [Mplus technical assistance discussion forum/thread]. Retrieved from http://www.statmodel.com/discussion/messages/9/425.html?1475502344

Muthén, L. K. (2012, January 12). Second order factors [Online discussion thread]. Retrieved from http://www.statmodel.com/discussion/messages/9/425.html?1461589742

Muthén, L. K. and Muthén, B. O. (1998-2010). *Mplus user's guide* (6th Ed.). Los Angeles, CA: Muthén & Muthén.

Muthén, L. K. & Muthén, B. (2009) Exploratory *factor analysis, confirmatory factor analysis, and structural equation modeling for continuous outcomes* [PowerPoint Slides].

Retrieved from Videos and Handouts for Mplus short courses: https://www.statmodel.com/course_materials.shtml

Muthén, B., & Muthén, B. O. (2009). *Statistical analysis with latent variables*. Wiley.

Netemeyer, R. G., Bearden, W. O., & Sharma, S. (2003). *Scaling procedures: Issues and applications*. Thousand Oaks, CA: Sage Publications.

Newmann, F. M., Marks, H. M., & Gamoran, A. (1996). Authentic pedagogy and student performance. *American Journal of Education, 104*(4), 280-312.

Newmann, F. M, Wehlage, G. G., & Lamborn, S. D. (1992). The significance and sources of student engagement. In F. Newmann (Ed.), *Student engagement and achievement in American secondary schools* (pp. 11–39). New York: Teachers College Press.

Nimon, K. F. (2012). Statistical assumptions of substantive analyses across the general linear model: a mini-review. *Frontiers in psychology*, *3*(322), 1-5.

O'Brien, R. M. (2007). A caution regarding rules of thumb for variance inflation. *Quality & Quantity, 41*(5), 673-690.

Oseguera, L., Locks, A. M., Vega, I. I. (2008). Increasing Latina/o students' baccalaureate attainment: A focus on retention. *Journal of Hispanic Education, 8*(1), 23-53. doi: 10.1177/1538192708326997

Pascarella, E. T. (1980). Student-faculty informal contact and college outcomes. *Review of Educational* Research, 50(4), 545-595.

Pascarella, E. T., Seifert, T. A., & Blaich, C. (2010). How effective are the NSSE benchmarks in predicting important educational outcomes? Change: The Magazine of Higher Learning, 42(1), 16-22.

Pascarella, E. T., & Terenzini, P. T. (2005). *How college affects students: A third decade of research* (Vol. 2). San Francisco, CA: Josey-Bass.

Patrick, H., Ryan, A. M., & Kaplan, A. (2007). Early adolescents' perceptions of the classroom social environment, motivational beliefs, and engagement. *Journal of Educational Psychology, 99*(1), 83-98. doi: 10.1037/0022-0663.99.1.83

Preacher, K. J. & MacCallum (2003). Repairing Tom Swift's electric factor analysis machine. *Understanding Statistics, 2*(1), 13-43.

Reinard, J. C. (2006). *Communication research statistics.* Thousand Oaks, CA: Sage Publications.

Reio, T. G., & Shuck, B. (2014). Exploratory factor analysis implications for theory, research, and practice. *Advances in Developing Human Resources*, 1-14. doi: 10.1177/1523422314559804

Reise, S. P., Waller, N. G., and Comrey, A. L. (2000) Factor analysis and scale revision. *Psychological Assessment*, 12(3), 287-297. doi: 10.1037//1040-3590.12.3.287

Robinson, C. C., & Hullinger, H. (2008). New benchmarks in higher education: Student engagement in online learning. *Journal of Education for Business, 84*(2), 101-109.

Rovai, A. P., & Barnum, K. T. (2003). On-line course effectiveness: An analysis of student interactions and perceptions of learning. *Journal of Distance Education, 18*(1), 57-73.

Ryan, A. M., & Patrick, H. (2001). The classroom social environment and changes in adolescents' motivation and engagement during middle school. *American Educational Research Journal*, *38*(2), 437-460. doi: 10.3102/00028312038002437

Sanders, R. L. (2006). The "imponderable bloom": Reconsidering the role of technology in education. *Innovate: Journal of Online Education, 2*(6).

Schmitt, N., & Kuljanin, G. (2008). Measurement invariance: Review of practice and implications. *Human Resource Management Review, 18*(4), 210-222. doi: 10.1016/j.hrmr.2008.03.003

Shaffer, D. R. (2005). *Social and personality development* (5th Ed.). Belmont, CA: Thomson Wadsworth.

Siegler, R. S., & Alilbali, M. W. (2005*). Children's thinking* (4th Ed.). Upper Saddle River, NJ: Prentice.

Skinner, E. A., & Belmont, M. J. (1993). Motivation in the classroom: Reciprocal effects of teacher behavior and student engagement across the school year. *Journal of Educational Psychology*, *85*(4), 571.

Skinner, E. A., Kindermann, T. A., & Furrer, C. J. (2009). A motivational perspective on engagement and disaffection: Conceptualization and assessment of children's behavioral and emotional participation in academic activities in the classroom. *Educational and Psychological Measurement, 69*(3), 493-525. doi: 10.1177/0013164408323233

Snyder, T. D., & Dillow, S. A. (2015). *Digest of Education Statistics 2013* (NCES 2015-011). National Center for Education Statistics. Institute of Education Sciences, U.S. Department of Education. Washington, D.C.

Strayhorn, T. L. (2008). The role of supportive relationships in facilitating African-American males' success in college. *NASPA Journal, 45*(1), 26-48.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics.* Saddle River, NJ: Pearson

Education.

Thayer, P. B. (2000). *Retention of students from first generation and low income backgrounds*. Washington, DC: National TRIO Clearinghouse.

Tinto, V. (1975). Dropout from higher education. A theoretical synthesis of recent research. *Review of Educational Research, 45*(1), 89-125.

Tinto, V. (1988). Stages of student departure: Reflections on the longitudinal character of student leaving. *The Journal of Higher Education, 59*(4), 438-455.

Tu, C. H., & McIsaac, M. (2002). The relationship of social presence and interaction in online classes. *The American Journal of Distance Education*, *16*(3), 131-150.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational research methods*, *3*(1), 4-70.

Warner, R. M. (2008). *Applied statistics: From bivariate through multivariate techniques.* Thousand Oaks, CA: Sage Publications.

Widaman, K. F., & Reise, S. P. (1997). Exploring the measurement invariance of psychological instruments: Applications in the substance use domain. *The science of prevention: Methodological advances from alcohol and substance abuse research*, 281-324. doi: 10.1037/10222-009

U.S. Department of Education, National Center for Education Statistics, Integrated Postsecondary Education Data System [IPEDS], Spring 2008 through Spring 2014, Enrollment component; and IPEDS Fall 2006 through Fall 2012, Institutional Characteristics component. (This table was prepared December 2014)

U.S. Department of Education, National Center for Education Statistics, Integrated Postsecondary Education Data System [IPEDS], Spring 2013 and Spring 2014, Enrollment component. (This table was prepared December 2014)

U.S. Department of Labor, Bureau of Labor Statistics [BLS]. (2016). *Employment Projections: Earnings and unemployment rates by educational attainment, 2015*. Retrieved from http://www.bls.gov/emp/ep_chart_001.htm

Van de Schoot, R., Lugtig, P., & Hox, J. (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology*, *9*(4), 486-492.

Zimmerman, B. J. (1990). Self-regulated learning and academic achievement: An overview. *Educational Psychologist, 25*(1), 3-17.